From the Department of Medical Epidemiology and Biostatistics
Karolinska Institutet, Stockholm, Sweden

MOLECULAR AND PSYCHOSOCIAL RISK FACTORS FOR CARDIOVASCULAR DISEASE

Iffat Rahman

Karolinska Institutet

Stockholm 2013

Published by Karolinska Institutet. Printed by Larserics Digital Print AB

ISBN 978-91-7549-112-7

To my parents.

Abstract

Cardiovascular disease (CVD) is the leading cause of mortality globally, and a major contributor to disability. There exist several well-established CVD risk factors, many of which are used in clinical practice. Nonetheless, these risk factors do not fully explain why certain individuals develop CVD. Several additional risk factors for CVD have been proposed which deserve to be examined further in prospective studies. Therefore, the overall aim of this thesis was to gain a comprehensive understanding of the epidemiology of well-established and promising risk factors for CVD.

In study I, we estimated the additive and non-additive genetic components contributing to variation in established CVD biomarkers. We could show that all of the traits were to some extent influenced by genetics, and that many of them were under the influence of non-additive genetic effects.

In study II, we examined how variation in anti-PC levels and Lp-PLA_2 activity is explained by genetic and environmental effects and how these effects are shared with other established CVD biomarkers. Both of these traits were found to be affected by genetic and environmental effects, Lp-PLA_2 activity was moderately correlated with several of the other biomarkers while anti-PC appeared to be regulated independently of more established CVD biomarkers.

In study III, we investigated whether clinical depression and use of antidepressants are associated with CVD outcome. Further, we examined if the associations were more specific for CHD or ischemic stroke. Depression was found to be a possible risk factor for the development of CVD, more specifically stroke.

In study IV, we investigated if individuals with any record of clinical depression or self-reported depressive symptoms had an increased risk for incident stroke after adjusting for a range of stroke risk factors. The association between depression and stroke could not be accounted for by traditional stroke risk factors.

In conclusion, CVD is a highly complex disorder affected by a multitude of risk factors, which in themselves are influenced by both our genetic make-up and environmental exposure. Although there exist well-established CVD risk factors useful in CVD risk assessment, novel CVD risk factors should be more thoroughly investigated in future studies. Such studies might not only add information that would be useful in CVD risk stratification, they could also enhance our biological understanding of this complex disorder.

List of publications

This thesis is based on the following studies, which will be referred to in the text by their Roman numbers (I-IV)

I. Genetic dominance influences blood biomarker levels in a sample of 12,000 Swedish elderly twins. *Twin Res Hum Genet. 2009 Jun;12(3):286-94.*

II. Genetic and environmental regulation of inflammatory CVD biomarkers Lp-PLA2 and IgM anti-PC. *Atherosclerosis. 2011 Sep;218(1):117-22.*

III. Clinical depression, antidepressant use and risk of future cardiovascular disease. *Resubmitted*

IV. Prospective study of clinical depression, self-reported depressive symptoms and their association with future stroke. *Manuscript*

List of Abbreviations

A	Additive genetic component
ApoA1	Apolipoprotein A1
ApoB	Apolipoprotein B
Anti-PC	Anti-phosphorylcholine IgM
BMI	Body mass index
C	Shared environmental component
CHD	Coronary heart disease
CI	Confidence Interval
CVD	Cardiovascular disease
CRP	C-reactive protein
D	Dominant genetic component
DNA	Deoxyribonucleic acid
DZ	Dizygotic
E	Unique environmental component
HbA1c	Hemoglobin A1c
HDL	High density lipoprotein
HR	Hazard ratio
ICD	International Classification of Diseases
LDL	Low density lipoprotein
Lp-PLA_2	Lipoprotein-associated phospholipase A_2
MI	Myocardial infarction
MZ	Monozygotic
NPR	National Patient Register
PDR	Prescribed Drug Registry
RNA	Ribonucleic acid

SALT	Screening Across the Lifespan Twin Study
STR	Swedish Twin Registry
TIA	Transient ischemic attack
TC	Total cholesterol
TG	Triglycerides

Table of Contents

1 Background

1.1 CVD and Atherosclerosis

Cardiovascular disease (CVD) is the major contributor to mortality and disability world-wide, as can been seen in figure 1. There has been a decline in CVD mortality in high-income countries over the years, but CVD mortality is unfortunately increasing in low- and middle-income countries [1].

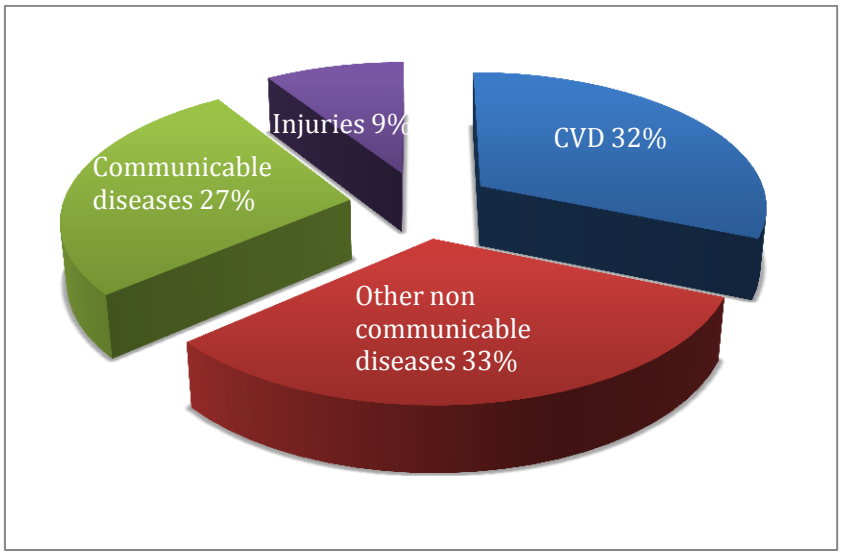


Figure 1. Chart showing the distribution of major causes of mortality worldwide in 2008 (adapted from the World Health Organization)

CVD include numerous complications which affect the heart and blood vessels, many of which have its origin in atherosclerosis. Atherosclerosis is a condition in which an artery wall is damaged due to plaque formation. The underlying mechanism involves the deposition and entrapment of modified LDL, for instance oxidized LDL, in the arterial wall which in turn attracts macrophages which internalizes the modified LDL upon which lipid peroxides are generated. Consequently cholesterol esters are accumulated and foam cells are formed [2-4]. Atherosclerotic lesions are characterized by a build-up of lipids in the artery walls, followed by inflammatory response.

Inflammatory and immune cells constitute an important part of an atherosclerotic lesion, the remainder being vascular endothelial and smooth-muscle cells [5]. When the endothelium, the single layer of cells on the arterial wall, is damaged, smooth muscle cells proliferate which leads to further narrowing of the arterial lumen. Macrophages and foam cells secrete metalloproteinases and tissue factors which degrade the vulnerable plaque which eventually leads to plaque rupture and thrombosis [2,6]. If the plaque rupture prevents blood flow through the coronary arteries it could lead to unstable angina or myocardial infarction [5,7]. Plaque rupture could also lead to stroke when the plaque obstructs a cerebral artery, or if the plaque originates elsewhere but detaches and moves to the circulation and occludes smaller vessels producing embolism [2].

1.2 CVD biomarkers

The cardiovascular research field has been very successful in finding biomarkers for CVD risk assessment. As an example, the framingham heart study has been a milestone in CVD risk stratification, investigating the relationship of total cholesterol (TC), low density lipoprotein (LDL), high density lipoprotein (HDL), age, and sex with 10 year risk of CHD development [8]. Besides LDL and HDL who constitute a major role in the development of atherosclerosis, their lipoprotein constituents' apolipoprotein B (apoB) and apolipoprotein A-I (apoA-I) have also shown to add information in CVD risk prediction [9].

A large number of studies have shown hyperglycemia, elevated triglycerides (TG) and glycated hemoglobin (HbA1c) to be associated with CVD risk, this association has been confirmed independent of diabetes. Elevated levels of TG are recognized as a risk factor in guidelines used for CVD risk assessment [10-13]. In a study comprising 29 Western prospective cohorts, which involved 262 525 study participants and a total of 10 158 incident CHD cases, triglycerides were reported to be moderately associated with an increased risk of CHD development [14]. The serum acute phase protein, C-reactive protein (CRP), has been proposed to be a risk marker for atherosclerotic progression, and has been suggested to be included in risk models for CVD screening [15,16].

Novel CVD biomarkers

The constituent papers of this thesis include investigations of the novel biomarkers Lipoprotein-associated phospholipase A_2 (Lp-PLA_2) and IgM antibodies against phosphorylcholine (anti-PC). These biomarker's influence on CVD development have not been as extensively studied as the more established CVD biomarkers mentioned earlier, and they are not used in clinical practice. Both these markers are indicators of atherosclerotic inflammation, in which Lp-PLA_2 has been suggested to exert both pro-inflammatory and anti-inflammatory effects, while anti-PC predominantly attenuates the inflammation in the vessel wall. They both target oxidized LDL and more specifically

its phospholipids component. Both of these biomarkers are described in more detail below.

Oxidative stress

Oxidative stress is one of the key mechanisms through which atherosclerosis is thought to develop. Reactive oxygen species are responsible for the process in which properties of phospholipids on lipoproteins are altered and hence become dysfunctional, immunogenic, and pro-atherogenic [17]. This is highlighted in the oxidation of LDL inside the arterial wall which elicits the inflammatory response, triggering the formation of atherosclerosis.

Lp-PLA$_2$

Lp-PLA$_2$also named plasma Platelet-activating-Factor (PAF)-acetylhydrolase is an enzyme secreted by macrophages and circulates primarily bound to LDL, and to much lesser extent to HDL. Although Lp-PLA$_2$ is expressed by inflammatory cells throughout the body, it is its production by atherosclerotic plaque which may elicit local stimulation of the innate immune system [18,19].

Lp-PLA$_2$ is capable of generating two bioactive pro-inflammatory mediators: free oxidized fatty acids and lysophosphatidylcholine [20,21]. However, since it also degrades PAF and oxidation products of phosphatidylcholine produced upon LDL oxidation and/or oxidative stress, it may also be considered as a potentially anti-inflammatory enzyme [19,22]. Nevertheless, a putative pro-atherogenic role has most widely been verified by earlier studies [23,24]. According to a comprehensive meta-analysis including approximately 80,000 individuals, Lp-PLA$_2$ activity could be found to increase the risk of future CHD and stroke even after adjusting for a range CVD risk factors [25]. Darapladib which is an inhibitor of Lp-PLA$_2$ activity has been found to decrease atherosclerotic plaque formation in ApoE-deficient mice as well as in pigs [26,27]. Further, a clinical study on patients with pre-existing coronary disease showed that treatment with darapladib in addition to statin treatment inhibited further growth of the necrotic core in the vulnerable plaque compared to placebo treatment [28].

Lp-PLA$_2$ has also previously been reported to be correlated with several other CVD biomarkers; LDL, HDL, apoB, TC, and TG, while weaker correlations were observed between Lp-PLA$_2$ and glucose, body mass index(BMI), and systolic blood pressure(SBP) [25,29,30].

Anti-PC

Phosphorylcholine (PC) is a major active component in bacteria, including *Streptococcus pneumonia*, and in apoptotic cells. It is also a component of oxidized LDL. Antibodies against PC (anti-PC) are thought to establish a first line defense against infections by *Streptococcus pneumonia* and maybe also other bacteria [31,32]. These antibodies belong to

the class of natural antibodies [33]. Scavenger receptors of macrophages bind oxidation-specific ligands, including phosphorylcholine-containing oxidized phospholipids, and consequently promote uptake of oxidized LDL [35].

A recent study demonstrated IgM anti-PC to inhibit uptake of oxidized LDL in macrophages, possibly reducing foam cell production [34]. Immunization with pneumococcal vaccine that induces IgM anti-PC antibodies has been demonstrated to be atheroprotective in mice and in hypertensive patients [35,36]. A previous study reported that low levels of anti-PC increase risk of ischemic stroke in men [37]. Moreover, it has been shown that levels of anti-PC are significantly higher in a population from New Guinea with a traditional lifestyle, as compared to Swedish controls, and that CVD is rare among individuals from Kitava, New Guinea [38].

1.3 Stroke

Stroke occurs due to disturbances of blood perfusion in the brain caused by either ischemia or hemorrhage. Ischemic stroke can be caused by for example embolism, local thrombosis, or systemic hypoperfusion (low blood perfusion throughout the body) [39]. Stroke diagnoses are based on clinical features and on data collected by tests such as computed tomography (CT), magnetic resonance imaging (MRI), cardiac imaging, duplex imaging of extracranial arteries, arteriography, and laboratory assessments. One of the most popular classification systems developed for classifying acute ischemic stroke is the Trial of Org 10172 in Acute Stroke Treatment (TOAST). The TOAST classification is based on clinical features as well as neurological and laboratory assessments. The system includes five categories; 1) large-artery atherosclerosis, 2) cardioembolism, 3) small-artery occlusion (lacune), 4) stroke of other determined etiology, and 5) stroke of undetermined etiology [40].

A stroke is diagnosed if a patient has a cerebral dysfunction with symptoms lasting for more than 24 hours or that leads to death [41]. Transient ischemic attack (TIA) refers to ischemic cerebral dysfunction that lasts less than 24 hours and is followed by complete neurological recovery [39].

Risk factors for stroke

According to the guidelines for stroke prevention developed by the American Heart Association there are several non-modifiable and modifiable risk factors for stroke. Well-documented non-modifiable risk factors include age, sex, ethnicity, genetic predisposition and low birth weight. Well-documented modifiable risk factors comprise of high blood pressure, atrial fibrillation, diabetes, cigarette smoking, physical inactivity, dyslipidemia, carotid artery disease, other CVD, sickle cell disease, postmenopausal hormone therapy, poor nutrition and obesity. Among less well-documented or potentially modifiable risk factors they list alcohol consumption, migraine, inflammation etc. [42].

A recent large-scale international study suggested that hypertension, cigarette smoking, diabetes, waist-to-hip ratio, poor nutrition, alcohol consumption, psychosocial stress, depression, previous cardiac disease, and ratio of apoB and apoA-I account for 90% of the risk (population attributable risk) of stroke [43].

1.4 The genome

The human genome resides in the nucleus of the cell and are packed into 23 chromosomes, although a small quantitative of the genome can be found in the mitochondria. The genome comprises of approximately 2,9 billion nucleotides, it can be roughly divided into coding genome (genes) and non-coding genome. The coding genome (genes) contains codes which are to be transcribed and translated into peptides or proteins, the coding genome contribute to approximately 1.5 % of the entire genome and the current estimate of protein coding genes in humans is 20 687. The non-coding genome contain genetic segments of diverse functionalities, these include noncoding RNA (e.g. transfer-RNA and ribosomal RNA), pseudogenes, introns, untranslated regions of mRNA, regulatory DNA sequences, microRNA, repetitive DNA sequences, and sequences related to mobile genetic elements. The vast majority (80.4%) of the human genome is believed to participate in at least one biochemical event in at least one cell type [44,45].

The human genome is not completely static throughout life but is to some extent plastic. Somatic mutations (mosaicism) do occur and the genome consists of mobile elements [46]. Mobile elements (transposons) are DNA sequences that have the ability to integrate into the genome at a new site within their cell of origin [45,47], however most transposons are inactive, only less than 0.05% of all transposons are considered to be active and can jump from one genomic region to another.

It is important to note that genes are not solely responsible for phenotypic variation and evolution, the non-coding genome also plays an important role. Indeed, the majority (above 90%) of trait-associated variants emerging from genetic association studies resides within non-coding genomic sequences. Much of the results points towards the involvement of genetic variants in transcriptional regulatory mechanisms, including variation of promoter and enhancer elements [48]. Furthermore, both microRNAs and retrotransposons have been suggested to be involved in complex biological processes such as human brain development and evolution [49,50].

Variation in genome can be measured by single nucleotide polymorphisms, insertions, deletions, copy number variations and chromosomal rearrangements. Genetic variations do however not necessarily result in a change in protein or in gene regulation [51-53]. In order to give a big picture overview of genetic variation examples of genomic comparisons between species and within species are presented further on in this section. The closest genetic relatives of humans are the bonobo and the chimpanzee. On

average, the genetic regions (sex chromosomes not included) in the bonobo genome have been found to be approximately 98.7% identical to corresponding sequences in the human genome [54]. The sequence of the chimpanzee genome has roughly 95 % similarity with the human genome sequence [55]. Human to human genomic similarity has been estimated to be about 99.5% similar, in other words there is 0.5% dissimilarity between the human genome of one individual to another [56]. Genetic variation among humans is thus not large and most of the variation is located in the non-coding genome, besides all genetic variation does not necessarily contribute to phenotypic variation. However, it is of importance to discover the proportion of the genetic variation that does result in phenotypic variation. Identification of such genetic variants may promote greater understanding of biological pathways underlying development of various diseases.

1.5 Biology of depression

The studies III and IV pertaining to this PhD thesis investigated the relationship between depression and CVD development. In this section hypotheses on the etiology of depression and its link to CVD will be discussed. According to the World Health Organization, in the year 2004 depression was the third leading contributor to the disease burden worldwide [57]. In Sweden, the prevalence of depression is high, especially among elderly in which 12-15% are affected by depression [58]. The etiology of depression is idiopathic, the diagnosis is subjective and rests on documentation of a certain number of symptoms that significantly impair functioning. Accordingly, depression can be viewed as a condition consisting of various diseases of diverse causes [59]. Nevertheless, there are many proposed hypotheses on the etiology of depression, a number of them are briefly discussed in this section.

Monoamine hypothesis

One of the most known hypotheses regards disturbances in the system of monoamine transmissions in the brain. The major types of antidepressants (selective serotonin reuptake inhibitors and tricyclic antidepressants) act through increasing certain monoamine transmissions in the brain. The affected monoamines are serotonin and norepinephrine, it is thus thought that depletion of these monoamines increases the predisposition to depression [60]. However, it usually takes weeks of treatment before mood-enhancing effects of antidepressant are initiated and the remission rate of antidepressants is not high which bring uncertainty to the monoamine hypothesis.

Neuroendocrinology

Hypothalamic-pituitary-adrenal (HPA) axis is a neuroendocrine system which controls the secretion of several hormones and consequently many biological processes, it is well-known for being involved in stress reactions [61]. Glucocorticoid levels, which are regulated by the HPA axis, have been shown to be raised in depressed patients. Physical and psychological stress increase levels of serum glucocorticoids. Depression has been

linked to metabolic disturbances such as diabetes, increased glucocorticoid levels can induce insulin resistance. Patients with Cushing's syndrome have abnormal cortisol levels and they display depressive symptoms [62].

Immune system

In animal studies cytokines have been reported to influence depressive like behavior [63].Numerous studies have demonstrated depressed patients to have activated inflammatory response displayed through higher levels of cytokines, CRP, platelet activation molecules and adhesion molecules [64]. But some studies have failed to establish associations between inflammatory markers and depression [65,66].

Vascular depression and poststroke depression

Vascular risk factors (such as hypertension and metabolic syndrome) and cerebral small vessel disease (infarcts and white matter lesion) have been shown to be associated with depression, more specifically depression among the elderly. Due to these findings the term "vascular depression" was coined which describes a subtype of depression affecting elderly individuals with vascular disruptions [67-69]. A systematic review investigating the association between vascular complications and late life depression could lend support to the vascular depression hypothesis, however only studies with cross sectional design were included in the meta-analysis [70]. It is important to stress that some studies on the depression-stroke link have found the association to be larger among those with early onset depression [71,72] (as opposed to late life depression). It could also be worth mentioning that this opposite trend could also be detected when investigating the study population used in study III pertaining to this thesis (supplementary tables S1 and S2).

Many observational studies have reported an association of depression occurring after stroke onset, this syndrome is called poststroke depression. It is expected that around a third of stroke patients will suffer from subsequent depression [73]. It has been suggested that poststroke depression is caused by lesions in certain brain regions [74]. Poststroke depression thus lends support to the notion of vascular depression.

The link between depression and CVD

Depression has been found to be associated with both CHD and stroke development according to two comprehensive meta-analyses [75,76]. It is not clear if depression in itself causes CVD or if the association is explained by other risk factors. A previous study found the association to be explained by behavioral mediators, namely physical activity, smoking and medication adherence [77]. Other explanations that have been suggested include immunological mechanisms, mechanisms of stress response, excessive platelet activity and toxicity of antidepressants [78-81]. A modest shared genetic vulnerability between major depression and coronary artery disease status has previously been reported [82]. Antidepressants use and its relationship with CVD risk has been examined

in several former studies, and the results have been conflicting. Some studies have shown that use of antidepressants increases the risk of CVD, even after adjusting for depression [83-86], and some have reported the antidepressants to be CVD protective [87-90], while some studies have reported null findings [91,92].

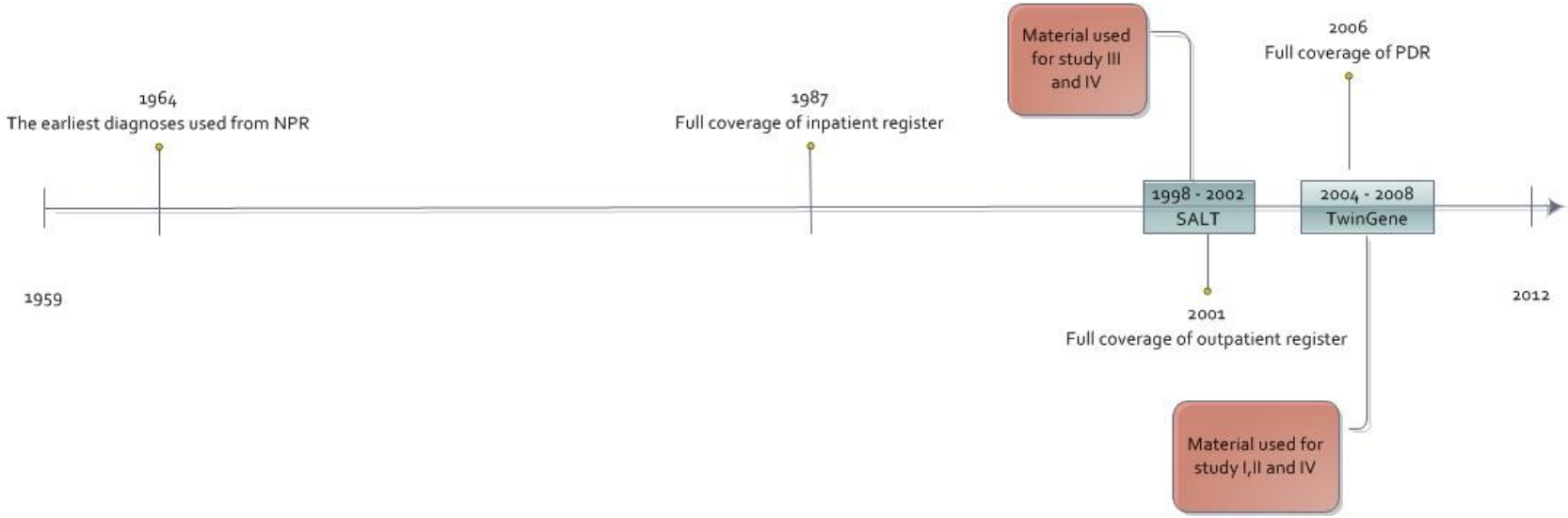
2 Aims of the thesis

The overall aim was to gain a comprehensive understanding of the epidemiology of well- established and promising risk factors for cardiovascular disease. Both genetic and environmental contribution to cardiovascular disease was considered.

More specifically, the objectives of the four separate studies were as follows:

- *Study I:* To estimate the variance components of HDL, LDL, total cholesterol, apoA-I, apoB, triglycerides, glucose, HbA1c, and CRP levels.
- *Study II:* To investigate to what degree the levels of the novel CVD biomarkers Lp-PLA_2 and anti-PC are affected by genetic and environmental factors
- *Study III:* To determine if clinical depression diagnosis and use of antidepressants are associated with CVD morbidity and mortality. Further, to examine if the associations are more specific for CHD or ischemic stroke.
- *Study IV:* To assess the association between depression and stroke after controlling for well-established stroke risk factors

Figure 2. Timeline depicting the collection of data and the establishments of cohorts and registers over the past decades. These are the registers and cohorts that have been the source of data for the four constituent papers of the PhD thesis.



3 Overview of study populations and registers

3.1 Swedish Twin Registry

The Swedish Twin Registry (STR) is one of the world's largest twin resources, currently containing 194,842 twins born between 1886 and 2008. STR obtains information on twin births occurring in Sweden from the National Board of Health and Welfare. It was established in the 1950's with a primary focus on epidemiological studies on CVD and cancer. The registry consists of three birth cohorts, two of these cohorts (the birth cohorts 1886-1925 and 1926-1958) have been used as study material for this PhD project. STR receives regular updates from health registers including the national patient register, the medical birth register, the prescribed drug registry, and the causes of death register. The twins have also been contacted in several additional waves with requests to participate in questionnaire/interview studies covering a broad selection of exposures, behavior and medical information. For the majority of same-sex twins in the STR, zygosity has been determined based on self-reported childhood resemblance, however DNA-based zygosity is available for 13% of all same-sexed twins in STR. Tests of the validity of similarity-based zygosity assignment by looking at genetic markers have overall yielded an accuracy estimate of around 98% among adults [93-95].

SALT

The study participants were identified from the population-based Swedish Twin Registry and had all participated in a computer assisted telephone interview called SALT (Screening Across the Lifespan Twin Study) conducted between 1998 and 2002. All screening data were collected over the telephone by trained interviewers with adequate medical background. Information on lifestyle and habits including educational level and smoking habits was collected. Special emphasis was put on diagnostic items that could determine whether a twin was likely to have a disease (rather than simply asking the twin whether they have a disease) [93,95]. The 11-item CES-D was administered during the SALT interview, which is a screening tool recommended to be used to measure current depression among elderly. A list of common prescriptions and non-prescription medication use was also recorded. A total of 44 826 individuals born before 1959 participated in the interview. Self-reported weight and height was available from the interview and BMI could be derived from these variables.

TwinGene

Twingene is a large population-based study on Swedish elderly twins born between 1911 and 1958. The study participants had previously taken part in the SALT interview. To be included in TwinGene, both twins within a pair had to be alive. In total, 12,647 individuals participated by donating blood to the study, and by answering extensive questionnaires about life style and health between 2004 and 2008. A total of 22,390 twins were invited to the study, thus the overall response rate was around 56%. The

participants were asked to make an appointment at their local health-care facility on Monday to Thursday mornings (not the day before a national holiday), to ensure that their blood sample would reach the KI Biobank in Stockholm the following day by overnight mail. A range of CVD blood biomarker levels were assessed from the blood samples, including lipids and inflammatory markers. Sampling and clinical blood test procedures have been described elsewhere [96]. Twingene has been linked to the Swedish national patient register, the causes of death register, and the Swedish psychiatric registry. The study was approved by the regional ethical review board at Karolinska Institutet and all participants gave informed consent. Study participants' weight and height were measured at the local health care facility, BMI was derived from these two variables. Systolic blood pressure (SBP) and diastolic blood pressure (DBP) were assessed by taking two measurements after five minutes rest. Age had a pronounced effect on the propensity to participate with a peak in participation for individuals born between 1936 and 1940, and therefore likely to comprise of recent pensioners. Sex did not affect participation since participation rates among males and females were very similar [94].

3.2 National Patient Register

The national patient register (NPR) in this case represents the inpatient register, which includes all hospital admissions that entailed at least one overnight stay, and the outpatient register which includes diagnoses registered during non-private specialized care. The inpatient register was established in 1964 and has full coverage since 1987. Currently, the coverage of the inpatient register is nearly 100%, while for the outpatient register it is about 80%. In the outpatient register, data from the private care are missing but the coverage of data from public care is almost 100%. Surgical day care procedures have been reported to the national patient register from 1997 and onwards, and since 2001 it is mandatory for counties to report outpatient physician visits. Diagnoses in the national patient register are coded according to the international classification of disease (ICD) system. The recording of the hospital admissions include a primary diagnosis and up to eight additional diagnoses as well as surgical procedures. The Swedish personal identity number enables the linking of registries. [97]

According to validation studies, the positive predictive values for heart failure, stroke and myocardial infarction in the inpatient register are in the range of 69-100% . Unfortunately, no validation studies have been undertaken to assess the validity of unipolar depression diagnoses. An increasing number of patients have been treated at the outpatient care over time. This might have led to a decreased sensitivity of the inpatient register in recent years for some diseases [98]. It has been reported that at least in Stockholm county there has been a shift from inpatient care towards more outpatient care for psychiatric patients in the 1990's [99].

3.3 Causes of Death Register

The causes of death register contains information on all deaths among Swedish residents, it was established in 1961 and is updated every year. The causes of death are classified according to the ICD. The register includes information on the main underlying cause of death as well as contributing causes of death. In some cases the deaths reported do not have information available from the death certificate, and thus the death is reported without a cause of death, in 2011 1.8% of the deaths reported lacked information on the cause of death [100].

3.4 The Prescribed Drug Registry

The national prescribed drug registry (PDR) records all prescribed drugs dispensed at pharmacies all around Sweden. Drugs that have been prescribed at hospital settings are not included in the register. The registry contains information on the identity of the drug according to the Anatomical Therapeutic Chemical (ATC) Classification System, the amount of drug articles (packages), price, prescription date etc. [101]. The registry contains information on prescriptions on some regions in Sweden from June 2004, and has full, national coverage since around January 2006.

3.5 Swedish Psychiatric Registry

Swedish Psychiatric Registry covers all psychiatric hospital discharges between the years 1963 and 1983. This registry contains up to six recorded diagnoses according to ICD-8. It includes information on patients receiving psychiatric treatment at all psychiatric hospitals in Sweden as well as all in-patient psychiatric departments of general hospitals [102].

4 Overview of methods

4.1 Twin modeling

One of the most popular twin methodologies to be utilized in epidemiological research is variance partitioning (heritability analysis). It is a method used to disentangle the relative contribution of genetic and environmental factors underlying a trait variance (total phenotypic variance). The methodology makes use of the genetic correlation between twins. Monozygotic (MZ) twins share 100% of their segregating genes (genetic correlation equals 1) while dizygotic (DZ) twins share roughly 50% of their segregating genes (genetic correlation equals 0.5). Studies examining genetic data have both supported the MZ genetic correlation [103] and refuted it [104]. In addition, it is assumed that MZ twins as well as DZ twins share their intrauterine and upbringing household environment (the period in life in which they are reared together) to the same extent. This assumption is called the equal environment assumption. In short, any trait similarity within a twin pair measured by intra-pair trait correlations ($r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2\sum(y-\bar{y})^2}}$) is due to either genetic or shared environmental factors. Trait dissimilarities within a twin pair on the other hand are due to unique environmental factors [105].

Below are equations for calculating broad-sense heritability component (h^2), additive genetic component (a^2), dominant genetic component (d^2), shared environmental component (c^2) and unique environmental component (e^2) in twin studies.

$rMZ = a^2 + c^2$

$rDZ = a^2/2 + c^2$

$a^2 = 2(r(MZ) - r(DZ))$

$d^2 = 4(r(MZ) - r(DZ))$

$c^2 = rMZ - a^2$

$h^2 = a^2 + d^2$

$h^2 + c^2 + e^2 = 1$

$rMZ + e^2 = 1$

$e^2 = 1 - r(MZ)$

Additive genetic variation

Additive genetic effect means that the overall genetic variation contributing to the phenotypic variation is the sum of all the individual effects of multiple loci (genotypes). Under an additive genetic model, it is assumed that there is no interaction between the alleles within a locus or between loci [105]. However, according to Falconer & Mackay additive genetic variance should not rest on the assumption of purely additive gene effects (i.e. that the genes act additively with absolute absence of interactions) since additive genetic effects can actually only be measured by molecular data [106]

Non-additive genetic variation

Non-additive genetic variance may enable the measurement of the portion of phenotypic variance which is due to epistatic interactions (gene-gene interactions) and dominance deviations (interactions within locus). In other words, epistasis depicts situations in which the total genetic effect is not the sum of all the individual effects across multiple loci and dominance genetics means that the total effect of a locus does not equal the sum of the effects of the two alleles in that particular locus. It is very difficult to estimate epistasis in a classical twin setting, if the MZ correlation is much larger than the DZ correlation it could be indicative of epistasis [105].

Multivariate analysis

It is possible to conduct variance partitioning of multiple traits. i.e. to analyze if the covariance of two or more traits can be attributed to common genetic or environmental variance. In study II multivariate variance partitioning was carried out implementing the cholesky decomposition model, in addition to obtaining a bivariate heritability estimate, a genetic correlation as well as unique environmental correlation was calculated. The *genetic correlation* indicates the extent to which genetic influences in one phenotype overlap with those of another phenotype. The genetic correlation could imply pleotropic genetic regulation contributing to multiple phenotypes. The bivariate heritability, also referred to as the *standardized genetic covariance*, reflects the genetic component of the total phenotypic correlation between two traits. The bivariate variance component estimates are achieved through calculating intra-individual within-trait correlations and comparing them to cross-twin cross-trait correlations [105].

Key assumptions in twin studies

Equal environment assumption

It is debated whether indeed MZ twins and DZ twins share similar degree of shared environment (intrauterine and upbringing household environment). MZ twins are often monochorionic twins i.e. they share the same placenta while DZ twins almost always have separate placentas and are thus dichorionic twins. There can also be competition between the twins in-utero and this could potentially affect the MZ twins more strongly

[107]. MZ twins might spend more time together compared to DZ twins later in life, this was investigated in study I [96]. Although many studies have pointed out differences in levels of shared environment for MZ and DZ twins, none of these differences have shown to have any substantial effect on heritability estimates.

Gene-environment interaction

It has been shown that gene expression variation can depend on environmental conditions [108], as such it could be plausible that individuals with similar genotypes who have been exposed to different environments will exhibit differences in the affected phenotype. But the variance due to gene-environment interactions is thought to be included in the environmental variance component [106].

Assortative mating

When individuals who resemble each other regarding certain phenotypes (for instance physical appearance) mate it results in assortative mating [107]. If the assortative mating (phenotypic similarity) is due to genetic similarity than the offspring of this mating might become more genetically similar. As a consequence the genetic correlation between DZ twins might be higher than 0.5, and thus one of the fundamental assumptions in twin modeling will be violated.

Direct effects in multivariate analyses

There is a concern regarding the cholesky model that it relies on the assumption that there is no direct effect of one phenotype to the other (meaning the effect is not due to shared genetic or environmental factors). If this assumption is violated both the genetic correlation and the environmental correlation will be inflated [109].

4.2 Linear regression

Linear regression is applied when dealing with a continuous outcome variable (y) is expected to be dependent on one or several exposure variables (x). The outcome variable needs to have a normal distribution, with an expected value $\beta_0 + \beta_1 x$ and variance σ^2. The equation of a simple linear regression line is

$$y = \beta_0 + \beta_1 x + e$$

where regression parameters

$$\beta_1 = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sum(x - \bar{x})^2}$$

$$\beta_0 = \bar{y} - \beta_1 \bar{x}$$

and e is the error term

Linear regression rest on the assumptions that the errors terms have the same variance for all observations, and that for any level of x, y is normally distributed [110].

4.3 Cox proportional hazards regression

The cox proportional hazards model is used for time-to-event analysis. The hazard rate (how long time it takes for an event to occur) is measured in the exposed and non-exposed group and the hazard ratio (the ratio of the hazard rates) informs if the hazard rate of the event occurring is higher or lower in the exposed group in relation to the unexposed group. It is a semi-parametric model since the underlying function of the baseline hazard function is unknown and thus lacks a mean and variance. It uses the partial likelihood function [111]. The model relies on the proportional hazards assumption, the assumption necessitates that the hazard rate in one level of a covariate in the model should be a constant multiple of the corresponding hazard rate of that covariates baseline level over time.

Time varying covariates

In cox regression, partial likelihood is used, a part of the likelihood is estimated each time an event occurs. In a cox model which incorporates time varying covariates, you allow the status of the covariate to change between the events, i.e. it considers intra-individual variation over time. An individual's risk can thus change over time due to change in covariate status, the individual can for instance go from unexposed to exposed.

4.4 Clustered data

In statistical analyses one of the most fundamental assumptions is that the observations in a model are independent from each other. Hence, if the observations are to some extent correlated, which is the case when utilizing data from twin pairs, this crucial assumption is violated. When assuming that the model has more independent observations than what it really does the power can be overestimated leading to too narrow confidence intervals of the regression parameters. Different statistical techniques can deal with correlated data, and often when considering the interdependence of the observations the standard errors become larger.

Robust sandwich variance estimator

The robust sandwich variance estimator can be used to measure the standard errors when the errors do not have the same variance for all observations or when the observations are not independent from each other [112]. Hence it can be implemented when dealing with clustered data. The robust sandwich variance estimator allows for a flexible model approach since it can handle structure of covariance-variance matrices that are misspecified.

Generalized estimating equations

One popular method to handle clustered data is by utilization of generalized estimating equations (GEE), since it will not rely on assumptions about the distribution of the error terms of your model [113]. GEE is flexible, the model allows you to specify a link function and can handle distributions such as normal, binomial and poisson. It uses the quasi-likelihood function, however the parameters produced by a GEE model has almost the same precision as those produced by maximum likelihood models [111]. GEE can use the robust sandwich variance estimator to calculate the standard errors.

5 Study summaries

5.1 Study I

In this study we aimed to estimate additive and non-additive genetic influences for HDL, LDL, total cholesterol, apoA-I, apoB, triglycerides, glucose, HbA1c, and CRP in a large sample of participants belonging to the Swedish Twin Registry. This was the largest twin study to date for several of the investigated biomarkers, providing data with better power than has previously been possible. Furthermore, since all the biomarkers have been assessed at the same occasion of all participants, direct comparisons between the estimates were facilitated.

Materials and methods

Study material was obtained from the TwinGene project which is a population-based study of Swedish twins born between 1911-1958, who were contacted and tested between 2004 and 2008. Informed consent was obtained from all participants. The participants were asked to make an appointment at their local health-care facility. The participants were instructed to fast from 8 PM (20:00) the previous night. A total volume of 50 ml of blood was drawn from each individual by venipuncture.

Assessment of serum samples

Clinical blood chemistry assessments were performed by the Karolinska University Hospital Laboratory. Levels of HbA1c were measured by a high-liquid performance chromatography separation technique. Levels of the other biomarkers were determined by Synchron LX systems (Beckman Coulter).

LDL levels were derived through the friedewald formula [114]. If you have measurements of total cholesterol, HDL and triglycerides this can be estimated by the following formula HDL = Total cholesterol – LDL – 0.45*Triglycerides, this formula can only be applied if plasma triglyceride levels are below 4.52 mmol/L.

Statistical analyses

Data handling and calculation of descriptive statistics as well as correlation coefficients were performed in SAS version 9.1 (SAS Institute, Cary, NC, USA). A variance component maximum likelihood method was implemented for estimation of variance components for each phenotype, using the Mx statistical program [115]. Univariate twin analyses were conducted in which the variance of the adjusted phenotypic values was divided into additive genetic effects (A), dominant genetic effects (D), shared environmental effects (C), and unique environmental effects (E). Scripts downloaded from the GenomEUtwin Mx-script library (http://www.psy.vu.nl/mxbib/)

Results

General characteristics of the study population, stratified by gender are summarized in Table 1.

Table 1 *General characteristics of study participants.*

	Men	**Women**
N[a]	5327	6421
Age (mean ± sd)	66.6 ± 8.7	65.9 ± 8.9
Weight (Kg, mean ± sd)	81.3 ± 12	68.3 ± 12
BMI (Kg/m^2, mean ± sd)	26.2 ± 3.5	25.7 ± 4.3
Individuals receiving statins (N[a])	744 (14%)	601 (9%)
Individuals receiving fibrates (N[a])	9 (0.2%)	3 (0.05%)
MZ[b] (N[a])	1281 (44%)	1656 (56%)
SSDZ[c] (N[a])	1817 (43%)	2376 (57%)
OSDZ[d] (N[a])	2175 (48%)	2346 (52%)
Unknown zygosity (N[a])	54 (0.56%)	43 (0.44%)

[a]Number of individuals, [b]Monozygotic, [c]Same-sexed dizygotic, [d]Opposite-sexed dizygotic

Table 2 shows the variance component decomposition with 95% confidence intervals (CI) from the ADE and AE models for all phenotypes. Additive genetic variance component was estimated to 0.66 for HDL, 0.64 for apoA-I, 0.48 for apoB, 0.50 for total cholesterol and 0.54 for TG. For these traits the dominant genetic component was not significant. Both additive and dominant genetic component were significant for LDL, glucose, and HbA1c. For LDL, additive and dominant genetic effects were estimated to 0.35 and 0.18, respectively. For glucose the corresponding estimates were 0.22 and 0.31, while for HbA1c additive genetic component was 0.16 and dominant genetic component was 0.55. In the ADE model the additive genetic component for CRP was not statistically significant, 0.13 (95% CI 0.00-0.28) while the dominant genetic component was estimated to 0.30 (0.12-0.46). Effect of non-shared environment was significant for all phenotypic traits.

Table 2 *Parameter estimates with 95% confidence intervals for additive genetic (a^2), dominant genetic (d^2) and non-shared environmental (e^2) variance components of age and sex adjusted trait levels in AE and ADE models.*

Phenotype	Model	a^2 (95%CI)	d^2(95%CI)	e^2(95%CI)
HbA1c	ADE	0.16 (0.03-0.29)	0.55 (0.55-0.68)	0.29 (0.27-0.32)
apoA-I	AE	0.64 (0.64-0.67)	-	0.36(0.33-0.37)
apoB	AE	0.48 (0.44-0.51)	-	0.52 (0.49-0.56)
HDL	AE	0.66 (0.63-0.68)	-	0.34 (0.32-0.37)
LDL	ADE	0.35 (0.21 -0.49)	0.18 (0.03-0.33)	0.47 (0.44 - 0.51)
TG	AE	0.54 (0.51-0.57)	-	0.46 (0.43-0.49)
Total cholesterol	AE	0.50 (0.46-0.53)	-	0.50 (0.47-0.54))
Glucose	ADE	0.22 (0.08-0.35)	0.31 (0.16-0.46)	0.47 (0.44-0.56)
CRP	ADE	0.13 (0.00-0.28)	0.30 (0.12-0.46)	0.57 (0.53-0.62)

[a]Pearson's correlation coefficient for MZ twins, [b]Pearson's correlation coefficient for SSDZ twins, [c]Pearson's correlation coefficient for OSDZ twins.

Since violation of the assumption of equal shared environment between MZ and DZ (i.e. MZ twins are exposed to more trait-relevant shared-environmental influences than DZ) would mimic genetic dominance, we investigated if discrepancies in reported contact frequency (i.e. the frequency by which the twins in a pair met in person) between MZ and DZ twins were present. MZ twins reported greater contact frequency than DZ twins, mean contact level was 3.00 for MZ twins while it was 2.57 for DZ twins (t-test, $p<0.0001$). Next, we investigated if contact frequency also was correlated with similarity in trait levels by computing the rank-order correlation (Spearman) between contact frequency and the absolute intra-pair difference in adjusted trait-levels. None of the zygosity specific correlations of trait difference and twin contact frequency reached significance except for difference in HDL among MZ twin pairs ($r=-0.058$, $p=0.04$).

The age at separation from co-twin (i.e. time shared same household environment) is also a measure of the degree of shared-environmental influences. Results showed that mean age at separation was significantly higher for MZ than for DZ twins, 19.7 years and 18.4 years, respectively (t-test, $p<0.0001$). For each separate zygosity strata the relation between absolute intra-pair difference in adjusted trait levels and age at separation was insignificant for all traits except apoA-I in MZ twins ($r=-0.061$, $p=0.04$).

Discussion

Our results for additive genetic effects for HDL, apoA-I, total cholesterol, and TG are consistent with what have been demonstrated in previous publications [116-131]. In addition, the contribution of non-shared environment was significant for all traits, which is also in agreement with what has previously been found. Here, we show for the first time significant effects of genetic dominance for LDL, CRP, glucose, and HbA1c in a population based twin sample. The reason for the novel findings of dominant genetic effects may be because of the enhanced power of the large and homogenous sample in our study compared to previous ones, enabling us to detect weaker variance components underlying the phenotypic traits. The high age of the study participants might also have been a contributing factor, leading to decreased influences from shared familial environment. In the case of CRP, additive genetic effect was found insignificant. This should not be taken as evidence for an absence of influences from additive genes (which appears biologically implausible for CRP) but indicates insufficient statistical power or some source of bias. Since variance component estimates are specific to the studied population, it is important to bear in mind that the obtained results not necessarily are representative of other populations or ethnic groups.

There appears to be no influence from shared environmental factors in this population. In the classical twin design comparing variance/covariance structures in MZ and DZ twins reared together it is not possible to model the effect of shared environment and dominance genetics simultaneously. Therefore, both influences may co-exist but their influences are not estimable in the same model. Another source of bias could come from violations of the assumption of equal importance of shared environmental influences between MZ and DZ twins.

By using data on contact frequency and age at separation available for a majority of the study participants, we demonstrated that there was evidence for differences in amount of shared environment between the zygosity classes. Even if MZ twins report significantly higher contact frequency and higher age at separation compared to DZ twins, we only found weak evidence for this to have an impact on twin trait similarity and more so this was only shown for HDL and apoA-I.

Large scale genetic association studies have been conducted to identify genetic variants influencing blood lipid levels. A genome-wide association study by Teslovich et al could

find 95 genetic variants significantly associated with TC, HDL, LDL, or TG. Some of these genetic variants were also associated with coronary artery disease. However, the genetic variants had small effect sizes and could only explain around 25-30% of the genetic variance of each lipid trait [132].

5.2 Study II

The aim of this study was to examine how variation in anti-PC levels and Lp-PLA_2 activity is explained by genetic and environmental effects and how these effects are shared with other established CVD biomarkers.

Materials and methods

The TwinGene project was used as study material. In total, 12591 individuals participated by donating blood to the study, and by answering questionnaires about life style and health. Detailed procedures for blood sampling have been previously described [96].

Measurements of Lp-PLA_2 and anti-PC

Lp-PLA_2 activity has been measured in 1600 individuals. Lp-PLA_2 activity was measured from plasma stored at –80°C in 96-well plates. Samples were measured in duplicate. Pooled human EDTA plasma from 20 normolipidemic human subjects served as an internal standard for all measurements. Lp-PLA_2 activity is expressed in nmol of degraded PAF per min per ml of plasma. The within-assay variability was ≤ ± 5% % [133,134].

IgM anti-PC levels were measured in a subset of 2036 TwinGene participants using an indirect non-competitive enzyme immunoassay (CVDefine ®, Athera Biotechnologies AB, Stockholm, Sweden) according to the manufacturer's instructions. The IgM anti-PC levels were expressed as arbitrary units (U/ml) estimated from a six point calibrator curve containing IgM anti-PC levels ranging from 0 to 100 U/ml [34].

Data handling and calculation of descriptive statistics as well as correlation coefficients were performed in SAS version 9.2 (SAS Institute, Cary, NC, USA). The proc genmod procedure (which applies generalized estimating equations) in SAS was implemented to perform linear regressions. A variance component maximum likelihood method was implemented for estimation of variance components for each trait, using the Mx statistical program [115]. Univariate twin analyses were conducted in which the trait variance was divided into additive genetic effects (A), dominant genetic effects (D), shared environmental effects (C), and unique environmental effects (E). Bivariate Cholesky models were analyzed in Mx for correlated traits, to partition the phenotypic correlation into A, C and E.

Results

The general characteristics of the study sample and distributions of the phenotypes by zygosity group are described in Table 3.

Table 3 *General characteristics of the study population.*

	MZ		**SSDZ**		**OSDZ**	
Variable	N	Mean (std dev)	N	Mean (std dev)	N	Mean (std dev)
Age (years)	1034	78.7 (4.05)	542	81.5 (4.11)	460	79.3 (2.38)
Lp-PLA_2 (nmol/ml/min)	779	61.3 (20.8)	434	62.9 (23.5)	374	64.8 (24.5)
Anti-PC (U/ml)	1034	89.1 (150)	542	72.1 (118)	460	74.6 (110)
CRP (mg/L)	980	3.81 (5.98)	515	4.37 (8.62)	445	3.92 (5.87)
ApoA1 (g/L)	1018	1.58 (0.30)	537	1.59 (0.30)	456	1.57 (0.29)
ApoB (g/L)	1018	1.12 (0.26)	537	1.10 (0.25)	456	1.13 (0.27)
TC (mmol/L)	1018	5.68 (1.15)	537	5.60 (1.17)	456	5.70 (1.27)
HDL (mmol/L)	1018	1.39 (0.40)	537	1.41 (0.41)	456	1.39 (0.41)
LDL (mmol/L)	1018	3.68 (1.02)	534	3.58 (1.02)	453	3.67 (1.03)
TG (mmol/L)	1018	1.38 (0.69)	537	1.37 (0.66)	456	1.40 (0.73)
Glucose (mmol/L)	1018	5.68 (1.22)	537	5.77 (1.23)	456	5.76 (1.24)
HbA1c (%)	1017	4.95 (0.65)	534	4.93 (0.69)	455	4.99 (0.73)
BMI (kg/m^2)	1034	26.0 (3.71)	542	25.7 (3.86)	460	25.8 (3.90)
Weight (kg)	1034	71.9 (12.5)	542	70.2 (12.6)	460	72.9 (13.5)
WC (cm)	1032	91.6 (11.4)	541	91.1 (11.5)	456	92.7 (11.9)

Table 4 shows the variance component decomposition with 95% confidence intervals (CI) from the ACE and AE model for Lp-PLA_2. The AE model was favored by the principle of parsimony since the χ^2 test was not significant. According to AIC, the ADE model was

to be preferred over the ACE model for anti-PC. Therefore, ADE and DE models with 95% CI are reported. Influence from unique environment was significant ($p<0.05$) for both of the traits. Contribution of additive genetic component was found to be 0.34 for Lp-PLA_2, while the corresponding estimate for anti-PC was non-significant. Even though the DE model is to be preferred for the variance component decomposition of anti-PC, it is not biologically plausible that dominance genetics would be the sole source for the genetic contribution in this case. This is most likely a result from insufficient power, nevertheless, the dominance genetic effect is 0.40 for anti-PC. No statistically significant evidence for influences of shared-environment was obtained for either of the biomarkers.

Table 4 *Parameter estimates with 95% CI for additive genetic (a^2), shared environmental (c^2), dominant genetic (d^2) and unique environmental (e^2) variance components of age and sex adjusted Lp-PLA_2 and anti-PC levels.*

Phenotype	a^2 (95% CI)	c^2 (95% CI)	d^2 (95% CI)	e^2 (95% CI)	Mx Model
Lp-PLA_2	0.34 (0.07-0.45)	0.02 (0.00-0.22)	-	0.64 (0.55-0.75)	ACE
	0.37 (0.27-0.45)	-	-	0.63 (0.55-0.73)	AE
Anti-PC	0.05 (0.00-0.29)	-	0.34 (0.09-0.44)	0.61 (0.56-0.66)	ADE
	-	-	0.40 (0.34-0.44)	0.60 (0.56-0.66)	DE

N = the number of individuals, r = Pearson's correlation coefficient, MZ = monozygotic twin pairs, SSDZ = same-sexed dizygotic twins, OSDZ = opposite-sexed dizygotic twins.

Anti-PC levels were correlated with CRP, apoB, TC, HDL, LDL, and Lp-PLA_2, however the magnitudes of the correlation coefficients were not large enough ($r<0.2$) for further investigations by bivariate variance partitioning. ApoB, TC and LDL were the biomarkers most strongly correlated with Lp-PLA_2 activity ($r>0.2$) which motivated further attempts to disentangle the contributing components by bivariate analyses (data not shown). The genetic overlap (rG) between Lp-PLA_2 and the other traits (apoB, TC and LDL) was in the range of 0.39-0.46. The corresponding overlap of unique environmental factors affecting the traits (rE) varied between 0.44-0.49 (Table 5). The covariance component decomposition was relatively similar for all three comparisons. Around one third of the total phenotypic correlation between Lp-PLA_2 activity and the other trait levels appears to be explained by genetic factors.

Table 5 *Genetic and unique environmental correlations and bivariate variance components decomposition with 95% CI*

Phenotypes	r_G (95% CI)	r_E (95% CI)	Biv h^2 (95% CI)	Biv e^2 (95% CI)
Lp-PLA$_2$-ApoB	0.39 (0.20-0.76)	0.49(0.41- 0.56)	0.33 (0.16-0.48)	0.67 (0.52-0.84)
Lp-PLA$_2$-TC	0.45 (0.24-0.89)	0.44 (0.35-0.51)	0.35 (0.18-0.50)	0.65 (0.50-0.82)
Lp-PLA$_2$-LDL	0.46 (0.27-0.83)	0.47 (0.39- 0.55)	0.36 (0.21-0.50)	0.64 (0.50-0.79)

Biv h^2 = Bivariate additive genetic component
Biv e^2 = Bivariate unique environmental component

Discussion

A heritable component could be found for both Lp-PLA$_2$ and anti-PC. For Lp-PLA$_2$, 0.37 of the total variance in enzymatic activity could be attributed to genetic variance. The highest cross-trait correlation was found for Lp-PLA$_2$ and LDL, apoB and TC were also moderately correlated with Lp-PLA$_2$. Further dissection of the covariance between Lp-PLA$_2$ and LDL revealed a genetic co-regulation explaining 36% of the total phenotypic correlation. Thus, 64% of the phenotypic correlation observed is explained by other factors. Lipid-lowering drugs with known reducing effects on LDL levels could also give concomitant reduction in Lp-PLA$_2$ activity. Lipid-lowering drugs may hence represent one of the co-regulating environmental factors. A previous report demonstrated that atorvastatin significantly reduced Lp-PLA$_2$ activity compared with placebo, even after adjusting for LDL [135].

Genetic variants in the APOE/APOC1 region have been associated with TC, LDL and apoB, and have also been found to be significantly associated with Lp-PLA$_2$ activity [136,137]. A recent study showed that several genetic variants related to LDL levels in humans are also associated with Lp-PLA$_2$ activity [138]. A former study pointed out a genetic region contributing to the variance in both LDL level and Lp-PLA$_2$ activity by genome-wide linkage analyses in baboons corresponding to the genetic region 2p24.3-p23.2 in humans [139]. These findings suggest that there are genetic regions that could possibly harbor genetic variants exerting pleiotropic effects on Lp-PLA$_2$ activity and LDL, TC as well as apoB.

Our heritability estimate for Lp-PLA$_2$ activity is lower than previously reported. Two former studies on the heritability of Lp-PLA$_2$ activity estimated the genetic component to be 0.62 and 0.54, respectively [29,140]. In the first study, Guerra et al. utilized 60 nuclear families (n=240) looking at parent-offspring Lp-PLA$_2$ activity relationship to measure heritability. In the second study based on 54 twin pairs, a genetic estimate of 0.54 with a p-value of 0.066 was reported [140]. A possible explanation for lower heritability estimate may be due to imprecision caused by the smaller sample sizes in the previous studies.

Our study population consists of a much larger sample size, thus the precision of the parameter estimates is higher.

This was the first time heritability analysis was conducted for anti-PC. A genetic component of 0.40 was observed for anti-PC levels. For anti-PC, the cross-trait correlations were weak (the highest correlation of 0.08 was observed together with Lp-PLA_2, LDL and TC). This could indicate that anti-PC is an independent biomarker for CVD, with a regulation that differs from the other CVD biomarkers assessed in this study. It should be mentioned that antibodies generally have complex regulations and functions and can undergo alterations in properties due to immunoglobulin class switching [141].

5.3 Study III

In this study we aimed to determine if clinical depression diagnosis and use of antidepressants are associated with CVD morbidity and mortality in a large population-based cohort. Further, we examine if the associations were more specific for CHD or ischemic stroke.

Materials and methods

The study participants were identified from the population-based Swedish Twin Registry and had all participated in a computer assisted telephone interview called SALT (Screening Across the Lifespan Twin Study) conducted between 1998 and 2002. Information on lifestyle, habits and health was collected. A list of common prescriptions and non-prescription medication use was also recorded. In total, 36,654 of the SALT participants were alive and free of CVD at baseline and included in this study. SALT has been linked to the national patient register, the causes of death register, the Swedish psychiatric registry and the prescribed drug registry.

Information on exposures and outcomes

Information on depression and CVD diagnoses were obtained through linkage to the national patient register. The national patient register in this case represents the inpatient register, which includes all hospital admissions that entailed at least one overnight stay, and the outpatient register which includes diagnoses registered during non-private specialized care.

The following diagnoses were used to define clinical depression; mild, moderate, severe depression with or without psychosis; atypical depression; recurrent depressive disorder with or without psychosis; other recurrent depression; dysthymia and mixed depression and anxiety. The following ICD codes were used; (ICD-10; F32.0-3, F32.8-9, F33.0-4, F33.8-9, F34.1 and F42.1), (ICD-9; 296C-D, 296W, 298A, 300E, 309A-B, 311X), (ICD-8; 296.00, 298.00, 300.40-41, 790.20), and (ICD-7; 314.99). In total, 1080 individuals were identified as receiving at least one diagnosis of depression.

CVD events were identified through ICD and surgical codes. The following ICD codes were used to identify cases of CHD (MI and unstable angina), heart failure, ischemic stroke and TIA; (ICD-10; I200, I21-2, I50, I63-4, G45), (ICD-9; 410-1, 433-6, 428), (ICD-8; 410-1, 432-6, 427), (ICD-7; 420, 332, 454, 434). In all, 2044 of the participants had suffered at least one incident CVD event.

Antidepressant use was identified through at least one prescription of the atc-code "NO6A" from the prescribed drug registry. The registry contains information on prescriptions on some regions in Sweden from June 2004, and has full, national coverage since 1 January 2006. Information on antidepressant use was also available from self-reported responses gathered in the SALT interview. Combining both data sources, a total of 6287 individuals had been prescribed antidepressants.

Additional information

Self-reported weight and height was available from the interview and body mass index (BMI) was calculated for the purpose of being included as a covariate. Information on self-reported smoking status, educational level, diabetes and hypertension was also taken from the interview. Educational level was based on five categories (1= 0-6 educational years, 2 = 6.5-9 educational years, 3 = 9.5-12 educational years, 4 = 12.5-15 educational years, 5 ≥ 16 educational years).

Data handling and generation of descriptive statistics were performed in SAS version 9.3. Cox proportional hazards regression using robust sandwich estimators to correct for correlated data was carried out in STATA 12.1. Number of days since entry to the study was used as the underlying time scale. Depression and antidepressant use were modeled separately as exposure variables with CVD as the outcome, both of these variables were modeled as time varying covariates. Subsequently, statistical interaction between antidepressants and depression as well as depression and gender were analyzed. Separate survival analyses for CHD and ischemic stroke were also carried out. Further, we restricted the definition of depression to only include those who had received a depression diagnosis before 1996. Date of baseline was 1 January 2006. End of follow up was occurrence of first CVD event, death, or 31 December 2009, whichever occurred first, thus the maximum follow-up time was 4 years.

Covariates included in the fully adjusted model were; birth year, gender, smoking status (ever or never smoker), educational level, hypertension (within the 5 past years), diabetes and BMI.

Results

Depression and use of antidepressants were both significantly associated with CVD outcome. There was however evidence for statistical interaction between depression and antidepressant use (HR=0.65, p=0.036) but not between depression and gender (HR=0.88, p=0. 51). After adjusting for all covariates the interaction term between depression and antidepressants was nearly significant (HR=0.69, p=0.069). Therefore we subdivided the exposure groups to “depression only” for depressed patients not using antidepressants, “depression and antidepressant users” for depressed patients on antidepressants and “antidepressants users only” for antidepressant users without a depression diagnosis.

In the fully adjusted model, significant increased risk of CVD development was observed among the depressed only group as well as among the antidepressant only group. Among depressed patients who also used antidepressants, a modest and non-significant increased risk was observed (HR=1.21, p=0.12) (Table 6). Among those who have ever been depressed, the HR was estimated to be higher for those who have not used antidepressants compared to those who have used antidepressants.

Table 6 *Hazard ratios for the association between clinical depression, antidepressant use and CVD*

Model 1*			Model 2†		
Variable	HR (95% CI)	P-value	Variable	HR (95% CI)	P-value
Depression only	1.60 (1.19, 2.15)	0.002	Depression only	1.50 (1.11, 2.02)	0.008
Antidepressants only	1.23 (1.08, 1.40)	0.002	Antidepressants only	1.17 (1.03, 1.34)	0.017
Depression + Antidepressants	1.28 (1.0, 1.64)	0.046	Depression + Antidepressants	1.21 (0.95, 1.55)	0.12

*Adjusted for birth year and gender
†Adjusted for birth year, gender, smoking status, educational level, hypertension, diabetes and BMI.

The results from a proportional hazards model which fitted the associations between depression, antidepressants and CHD or ischemic stroke are depicted in Table 7. Depression was associated with a marked increase in risk of developing incident stroke regardless of exposure to antidepressant status, there was no significant association between ”antidepressants users only” and stroke. Neither depression nor antidepressant use were significantly associated with CHD (Table 7).

Table 7 *Hazard ratios for the association between depression, antidepressant use and CHD or stroke.*

CHD			Stroke		
Variable	HR (95% CI)*	P-value	Variable	HR (95% CI)*	P-value
Depression only	1.28 (0.80, 2.06)	0.28	Depression only	1.78 (1.16, 2.74)	0.009
Antidepressants only	0.99 (0.79, 1.23)	0.90	Antidepressants only	1.18 (0.96, 1.45)	0.13
Depression + Antidepressants	0.81 (0.52, 1.27)	0.36	Depression + Antidepressants	1.74 (1.26, 2.40)	0.001

*Adjusted for birth year, gender, smoking status, educational level, hypertension, diabetes and BMI

In order to limit the influence of reverse causality between depression and stroke we performed analyses in which we restricted the definition of depression to only include those who had received a depression diagnosis before 1996. (i.e. at least 10 years before baseline). In a fully adjusted model, the significant association between depression and stroke (HR=1.70, $p<0.01$) remained.

Discussion

This large prospective population-based study on elderly Swedish citizens shows that depression is an important risk factor for later development of CVD. The strongest association was observed for depression without use of antidepressants. This indicates that associations between antidepressants and CVD observed in in this study and previous studies could have been due to confounding by indication. Confounding by indication means that the indications for drug use (depression is an indication for use of antidepressants) could confound the drug-disease association so that it appears as if the drugs causes the disease [142]. Another important finding in this study is that the association with depression appeared to be specific to stroke and not CHD. We could also show that the association remained for stroke when only looking at depression diagnoses recorded before 1996.

Depression has been more widely studied in relation to CHD rather than stroke [75,79] but we found depression to be more strongly associated with stroke. A recent comprehensive meta-analysis found the association between depression and future

stroke to be significant [76]. However, two previous studies investigating depressive symptoms in relation to both CHD and stroke in the same study population had the opposite finding compared to ours [143,144]. The reason may have been due to dissimilarities in study characteristics, definition of exposure and outcome as well as exclusion criteria. None of the two aforementioned studies used clinical records to define depression. In the Nabi et al. study [143], all participants who had had a CHD or stroke event at baseline (1998) were excluded. However it was not stated if those who had another type o CVD event prior to baseline also were excluded, thus reverse causality could have been an issue in their study design. Wassertheil-Smoller et al. [144], conducted separate analyses for those with a CVD history and those without, using self-reported data for identifying prevalent cases. But the validity of self-reported CVD data might be questionable.

This study does not recommend the use of antidepressants as a proxy measurement for depression. It is important to point out that antidepressants are not exclusively used to treat depression, other conditions for which antidepressants are used include anxiety disorders, obsessive compulsive disorders, eating disorders, chronic pain, posttraumatic stress disorder, and social phobia [145,146].

We found a weaker association with CVD for depressed patients who did receive antidepressants compared to those we did not use antidepressants. Nevertheless, we cannot draw the conclusions that antidepressants indeed are CVD protective. It should be mentioned that the "depressed only" group in this study did not receive less medical drug prescriptions in general, they actually received more prescriptions for warfarin and antihypertensive drugs (beta blockers not included) compared to the other exposure groups (data not shown).

Depression and use of antidepressants are variables that change over time, and in order to reduce misclassification bias during the follow-up period, we modeled the exposures as time varying covariates. However the change in exposure status was only modeled when an unexposed study subject turned into an exposed subject, if the order was the reverse it was not taken into account in the model. Some of the other covariates such as BMI can also vary over time but they could not be modeled as time varying covariates due to lack of longitudinal data.

It is possible to adjust for "familial confounding" employing family-based designs, such as the co-twin control design. But, it would not be feasible to conduct such a method since both depression and antidepressants were modeled as time varying covariates. What more is, there would have been a significant loss o of power since only complete twin pairs could be included in such study designs. It is worth mentioning that the validity of the co-twin control design is based on several strong assumptions. For instance, measurement error of the exposure variable can easily distort the results, and there undoubtedly exists issues with misclassification of the exposure variables in our

cohort. Methodological issues with family-based designs have been discussed thoroughly in a previous paper [147].

It is important to point out that the study design does not permit inference on causality. We cannot conclude that clinical depression by itself causes CVD development. The observed association could for instance be the result of unmeasured confounders or residual confounding. Also, using self-reported data for diabetes, BMI, educational level and hypertension is far from optimal, actual measurements on blood pressure, height and weight would have been advantageous.

A big study limitation is that information from primary care is lacking. A majority of those who develop depression after 65 years only come in contact with primary care and are not admitted to specialized psychiatric care in Sweden. Hence, they do not get their depression diagnosis reported in the national patient register [58]. It is possible that the effects of depression on CVD and stroke risk were biased due to misclassification of exposure in those above 65 years of age.

The study suggests that individuals who at some point in life are clinically depressed should be monitored more closely for their increased risk of developing CVD, particularly ischemic stroke. Further studies are needed to confirm and gain a deeper understanding of the association between clinical depression and stroke.

5.4 Study IV

With the aim to bring further clarity to the relationship between depression and stroke, we here investigated if study participants with any record of clinical depression or self-reported depressive symptoms had increased risk for incident stroke morbidity and mortality after adjusting for a range of stroke risk factors.

Materials and methods

The study material was obtained from the Twingene study [94] and the SALT interview. In total, 12,647 individuals participated by donating blood to the TwinGene study, by undergoing a health examination, and by answering a questionnaire about lifestyle and health between May 2004 and 2008. Sampling and clinical blood test procedures have been described elsewhere [96]. Twingene has been linked to the Swedish national patient register, the causes of death register, and the Swedish psychiatric registry.

Information on depression diagnoses and CVD diagnoses were obtained through linkage to the national patient register (inpatient and outpatient register) and the Swedish psychiatric registry. Detailed description on clinical depression and stroke definition has been reported elsewhere [148].

Identification of Self-reported depression cases

The 11-item CES-D was administered during the SALT interview, which is a modified shorter version of the original 20-item CES-D. Shorter versions of the 20-item CES-D have been recommended to be used to screen for current depression among elderly, since study participants of old age have been shown to experience difficulties completing the longer version [149]. The 11-item CES-D consists of 11 questions regarding feelings and symptoms of depressive mood over the past week. Each question is scored from 0-3, thus the maximum score possible is 33. The recommended cut-point of 9 was used to define depressive mood [150].

Information on covariates

Information on use of antihypertensive drugs, lipid lowering therapy, presence of diabetes, presence of migraine and birth weight was obtained from self-reported questionnaires from TwinGene. Study participants' height and weight were measured at the local health care facility. Body mass index (BMI) was derived from these two variables. Systolic blood pressure (SBP) and diastolic blood pressure (DBP) were assessed by using the average of two measurements after five minutes of rest. Information on participant's low density lipoprotein (LDL), high density lipoprotein (HDL) and C-reactive protein (CRP) levels were taken from the TwinGene measurements. Information on smoking status and alcohol consumption (ever or never smoker/drinker) were taken from the SALT interview. Cases of atrial fibrillation were identified from the national patient register with the ICD codes: (ICD-10; I48), (ICD-9; 427D), (ICD-8; 427.92), and (ICD-7; 433.12-3).

Statistical analyses

Data handling and generation of descriptive statistics were performed in SAS version 9.3. Cox proportional hazards regression using sandwich estimators to correct for correlated data was carried out in STATA 12.1. The TwinGene study was conducted between 2004 and 2008 and the date of baseline varied accordingly. Number of days since baseline (entry to the TwinGene study) was used as the underlying time scale. All study participants who had ever been diagnosed with clinical depression before censoring were classified as exposed to depression. Clinical depression and atrial fibrillation were modeled as time-varying covariates. End of follow-up was 31 December 2010 rendering the maximum follow-up time to be 6.5 years. The primary outcome was stroke, including ischemic stroke, hemorrhagic stroke, and transient ischemic attack (TIA). Clinical depression and CES-D rated depression were modeled separately as exposure variables

Covariates included in the fully adjusted model were; age, sex, atrial fibrillation, BMI, HDL, self-reported diabetes, SBP, DBP, self-reported use of antihypertensive medication, lipid lowering therapy, alcohol intake and smoking status. Values for BMI were missing for 529 study participants, for 659 study participants for DBP and 658 individuals had missing information on SBP. For subjects missing information on SBP, DBP or BMI imputed values were obtained by using the average SBP, DBP or BMI of their corresponding age group (in 5 year intervals). Associations between LDL and CRP levels with stroke were also assessed, however since none of these biomarkers were associated with stroke outcome and had many missing values they were omitted from the fully adjusted models. In addition, birth weight and migraine were also analyzed in relation to stroke, however they were not associated with the outcome and also omitted from the final statistical models. The proportional hazards assumption was examined with Schoenfeld's residuals. If the assumption was violated by any of the covariates, that covariate was stratified in the proportional hazards model which enabled different baseline hazards for the different levels of the covariate.

Results

Descriptive characteristics by exposure and outcome group are depicted in Table 8. The proportion of stroke was higher among those who were clinically depressed compared to those who were not diagnosed with depression. Among those who had been diagnosed with depression only 29% had been classified as having CES-D depressive symptoms.

Table 8 *Descriptive characteristics by exposure or outcome category*

Variable (unit)	Non-depressed (SD)	Clinically Depressed (SD)	No stroke (SD)	Stroke (SD)
N	10941	397	10981	357
Age at study entry (years)	64 (7.9)	64 (7.7)	64 (7.9)	71 (7.7)
Clinical Depression (%)	-	-	3.5	6.4
CES-D score depression (%)	8.9	29.0	9.6	9.2
Stroke/TIA (%)	3.1	5.6	-	-

*Values are means and standard deviations (SD) or percentages at baseline except for clinical depression where information was also collected during follow up period.

Clinical depression was significantly associated with stroke, but CES-D rated depressive symptoms were not. The results from the crude model and the fully adjusted model for clinical depression and stroke were very similar, the hazard ratio (HR) was 2.24 in the crude model and 2.23 in the adjusted model (Table 9).

Table 9 *Association between depression and stroke*

		Model 1[a]		Model 2[b]	
Variable	No cases	HR (95% CI)	P-value	HR (95% CI)	P-value
All stroke/TIA	357				
Diagnosed depression		2.24 (1.44, 3.48)	<0.001	2.23[c] (1.42, 3.49)	0.001
CES-D score depression		1.19 (0.83, 1.70)	0.34	1.24[c] (0.86, 1.79)	0.25

[a] adjusted for age and sex

[b] adjusted for age, sex, BMI, SBP, DBP, HDL, diabetes, smoking, alcohol intake, antihypertensives, statins, and atrial fibrillation.

[c] Stratified for antihypertensives

Results from analyses specifically investigating the associations between clinical depression and ischemic stroke/TIA or hemorrhagic stroke are shown in Table 10. The results from the crude and fully adjusted models were very similar. The association was significant for ischemic stroke/TIA (HR=2.16, p=0.003) after adjusting for all covariates: The corresponding point estimate for hemorrhagic stroke was larger but non-significant (HR= 2.46, p=0.14).

Table 10 *Association between depression and ischemic stroke/TIA or hemorrhagic stroke*

		Model 1[a]		Model 2[b]	
Variable	No cases	HR (95% CI)	P-value	HR (95% CI)	P-value
Ischemic stroke/TIA	300				
Diagnosed depression		2.15 (1.31, 3.52)	0.003	2.16 (1.30, 3.56)	0.003
Hemorrhagic stroke	44				
Diagnosed depression		2.52 (0.78, 8.15)	0.12	2.46[c] (0.73, 8.27)	0.14

[a] adjusted for age and sex
[b] adjusted for age, sex, BMI, SBP, DBP, HDL, diabetes, smoking, alcohol intake, antihypertensives, statins, and atrial fibrillation.
[c] Stratified for dichotomized DBP

Discussion

In this study we could show that clinical depression is a prospective risk factor for stroke in individuals free of CVD at baseline even after adjusting for a battery of stroke risk factors. We could not find a significant association between self-reported depressive symptoms and stroke. Further, the association with clinical depression was similar in magnitude for stroke of both ischemic (including TIA) and hemorrhagic type, but only statistically significant for the former. Due to the smaller number of hemorrhagic stroke outcomes (N=44) it may just reflect lack of statistical power. It would be prudent to examine the relationship between depression and hemorrhagic stroke more rigorously in a study sample with a sufficient number of events.

To our knowledge this is the first study which has compared how diagnosed depressions obtained from patient registers and self-reported depressive symptoms are related to incident stroke in the same material. It is worth mentioning that these are two very different measurements of depression. The clinical diagnoses consisted of any event of unipolar depression of varying severity that could be detected in the national patient register or psychiatric registry from 1960's and onwards, the self-reported depressive symptoms were based on the 11-item CES-D administered during the SALT interview and examining depressive mood during the week before the interview.

From the meta-analysis by Pan et al. it is possible to distinguish between studies which have used clinical diagnoses on depression examined by a physician/psychologist and

self-reported depression diagnoses. By combining the meta-analysis [76] with the results from study III and study IV it is possible to obtain pooled estimates of the RR for clinical diagnoses and self-reported diagnoses separately (excluding studies using mixed clinical/self-reported definition of depression diagnoses [151]). Using a random effects model a pooled estimate of the relative risk (RR) found in studies that have used clinical diagnoses to define depression could be estimated to RR=2.49, CI 1.71-3.62 (test for heterogeneity, p <0.01). The corresponding pooled estimate for studies using self-reported depressive symptoms was found to be RR=1.32, CI 1.21-1.45 (test for heterogeneity, p=0.03). This could be indicative of a dose-response relationship.

We used a broad definition of unipolar depression, including dysthymia which is regarded as a milder chronic mood disorder separate from the cluster of major depressive disorder (MDD) subtypes [152]. This was necessary since ICD-7 and ICD-8 versions did not have different codes to distinguish between subtypes of clinical depression. A DSM-IV definition of MDD would have allowed for a more homogenous exposure variable. But a previous study which conducted separate analyses for MDD and dysthymia to investigate their associated risks with stroke outcome reported that the risk estimates were very similar[153]. In this study we attempted to compare CES-D self-reported measurement with clinical measurement of depression. It is thus important to note that the CES-D is used as a tool to measure "depressed mood" which is not specific to MDD. Moreover, subtypes of depression are defined according to the manifestation of symptoms, they are not based on distinct pathophysiological mechanisms [59].

A majority of those who develop depression after 65 years of age are not admitted to specialized psychiatric care in Sweden and therefore do not get their depression diagnosis reported in the national patient register [58], and the psychiatric registry covers psychiatric admissions only before 1983. Therefore, the association between depression and stroke reported in this study could have been biased.

We cannot conclude that the relationship between depression and stroke is causal, the observed association could for instance be due to unmeasured or residual confounding. Biological confounders/mediators suggested by earlier studies to contribute to the link between depression and stroke include increased platelet activation [80,154,155] which was unaccounted for in this study. There is also a possibility that depression might be a symptom of a prior silent brain infarction [156], and it has been reported that silent brain infarctions increases the risk for future stroke [157].

In conclusion, evidence showing that clinical depression is a contributor to stroke development is increasing. Biological mechanisms underlying the observed association between depression and stroke have yet to be unraveled. Prospective studies elucidating the biological mechanisms behind the relationship between depression and stroke are required.

6 General discussion

6.1 Methodological considerations

"A hair, they say, divides the False and True…"

Omar Khayyam (1048-1131), mathematician, astronomer, philosopher and poet.

In this section some very general methodological issues in epidemiological studies are brought up. Accuracy in population studies relies upon random errors (precision) and systematic errors (biases). Only systematic errors (which affect the validity of the study) will be addressed herein.

6.1.1 Internal validity (bias)

Selection bias

Selection biases arise due to factors influencing the selection of study subjects and study participation and lead to distortions of the study results [158]. The obvious source of selection bias in the study material used in this PhD thesis is that only twins were included, non-twin population was excluded. In addition, only elderly individuals of Swedish origin were selected. TwinGene might particularly be subject to selection bias, both twins in a twin pair must have survived at least until May 2004, and in order to participate the twins had to be healthy enough to visit a local health facility and donate blood. In TwinGene the proportion of clinically depressed individuals was considerably low (2.9%), while it was 5.1 % in SALT. What more is, the proportion of smokers in Twingene is lower than that of the general Swedish population [159]. It is possible that TwinGene might suffer from a "healthy volunteer effect", and consequently the risk estimates could to some degree be biased [160].

Selection bias and biased heritability

It has been reported that selection bias can inflate additive genetic and unique environmental components and attenuate shared-environmental component. High levels of selection bias could also result in spurious non-additive genetic components [161]. Heritability estimates will only be biased if the propensity to be included in a study is correlated with the phenotype of interest. When there is differential participation in relation to the phenotype of interest the estimate of the population correlation between twins will be reduced, and this effect will be larger for the lower correlation (the DZ correlation is often lower than the MZ correlation). As mentioned earlier, age had an effect on the tendency to participate with a peak in participation for subjects born between 1936 and 1940. The age window was even narrower in Study II [162], with mean age around 80 years old (standard deviation ~ 4 years), both Lp-PLA_2 and anti-PC are correlated with age. In addition, there were also other criteria for inclusion into the study which is listed in Study I [96]. It can be debated whether excluding observations of

CRP values above 10 mg/L before conducting the heritability analyses on CRP (higher CRP values reflect infection induced inflammation) was the best decision.

Confounding (mixing of effects)

Confounders are variables in a statistical model which are associated with both the exposure and outcome of interest, hence a spurious association between the exposure and outcome of interest might be observed solely due to the confounder if unaccounted for. Thus in statistical models assessing epidemiological relationships, variables suspected to be potential confounders need to be conditioned on [158]. Confounders and biased associations have particularly been discussed in studies III and IV.

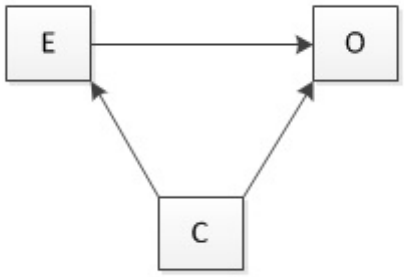


Figure 3. Direct acyclic graph delineating the relationship between an exposure (E) a confounder (C) and an outcome (O)

Colliders

In some instances an exposure and outcome of interest might both contribute to a third variable, a collider. As a hypothetical example a disease might either be caused by risk factor E (exposure of interest) or risk factor O (outcome of interest). Say that there is no association between these two risk factors, but we attempt to investigate the causal relationship between risk factor E and risk factor O by conditioning on the disease (the collider), this can give rise to a spurious negative association between risk factor E and O. The collider could also be a register (if the exposure and outcome status affects the likelihood of being included in the register) used as material for an observational study, in that case a selection bias (only selecting individuals included in the register) could introduce a spurious association between exposure and outcome [163].

Figure 4. Direct acyclic graph illustrating the relationship between an exposure (E) a collider (C) and an outcome (E)

Mediators

Mediators are variables in a statistical model which mediates the casual effect of the exposure on the outcome of interest. More specifically, the exposure variable causes a change in the level of the mediator variable which in turn affects the outcome variable. The mediator is said to be in the "causal pathway" between the exposure and outcome of interest. In studies III and IV it was difficult to distinguish between potential confounders and mediators.

Figure 5. Direct acyclic graph illustrating the relationship between an exposure (E) a mediator (M) and an outcome (E)

Dealing with mediators can be precarious since erroneously adjusting for a mediator can in fact give rise to a spurious association between the exposure and outcome. Such biased associations can occur when the association between the mediator and the outcome variable is confounded, in a direct acyclic graph these confounders would be termed as ancestors to colliders.

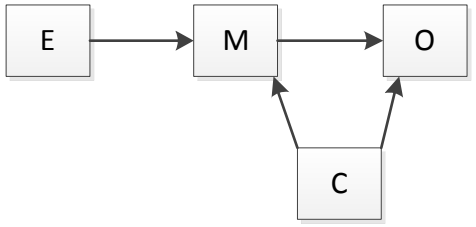


Figure 6. Direct acyclic graph delineating the relationship between an exposure (E) a perceived mediator (M) ancestor to a collider (C) and an outcome (O)

Reverse causality

Reverse causality denotes the chicken and the egg dilemma, i.e. what is perceived as the order of exposure-outcome relation in a study setting is actually reversed in reality. When the exposure variable is time varying and the outcome variable is hard to measure, or when molecular processes are cyclic (reciprocal causality), reverse causality can beset epidemiological studies. In study III, we attempted to reduce the potential for reverse causality, by restricting the exposure variable (depression

diagnosis) to be registered at least 10 years before the outcome (incident CVD diagnosis). However, reverse causality can also be a problem even when temporality, i.e. the exposure is measured at a time point preceding the measurement of the outcome, is included in the study design. The exposure and outcome variables in the study are usually proxy measurements of the true exposure and outcome, and it is not feasible to have access to data on the time varying exposure measured at every single time point throughout a study subject's life, the same applies for the outcome variable. Reverse causality also applies to mediators and confounders, in study IV it was difficult to determine whether blood pressure was a confounder or mediator in the association between depression and stroke. Depression diagnoses were usually reported before baseline while blood pressure measurements were taken at baseline, intuitively blood pressure could be perceived as a mediator. However, some depression diagnoses were reported at a time point after baseline. The association was only significant when assessing blood pressure and depression where blood pressure measurement was preceding the depression diagnoses, the association was not significant when the blood pressure measurement succeeded the depression diagnoses (Supplementary table S3). Given those results, it becomes evident that blood pressure could possibly have exerted a confounding effect. Owing to reverse causality it can be very difficult to discern between confounders, mediators and colliders.

Information bias (Misclassification bias)

In this section varieties of misclassification bias (measurement error of discrete variables) most likely to have been affecting the studies belonging to this thesis will be discussed. Misclassification bias simply refers to measurement errors of the exposure variable, the outcome variable or any of the other covariates included in a model of interest. Non-differential misclassification of exposure refers to the situation when there is a misclassification of the exposure variable which is not dependent on the other variables in the model. Non-differential exposure misclassification is regarded to weaken the association between exposure and outcome, i.e. lead to an underestimation of the true effect, although this notion has been challenged by simulation studies since it does not consider other conditions such as random error [164]. Non-differential misclassification of a dichotomous confounding variable will not fully adjust for that particular confounder and thus the association between the exposure and outcome will remain biased.

Differential misclassification bias on the other hand arises when misclassification of exposure is dependent of other variables in the model. Differential misclassification of exposure can either lead to either an underestimation or overestimation of the true effect. Studies III and IV were afflicted by differential misclassification of clinical depression depending on age, study participants above 65 years of age were more likely to be misclassified as non-depressed. Age is positively associated with stroke outcome, if assuming that the pathophysiology of early-onset and late-onset depression do not

differ the misclassification of depression is likely to have resulted in an underestimation of the depression-stroke association.

6.1.2 External validity

Generalizability

The population used in an epidemiological study is a sample from a target population [158]. Generalizability refers to how representative the study population is of the target population. There are concerns regarding the generalizability of the study population utilized in the constituent studies of my thesis. Our study population consists of Swedish twins and twins might differ from singletons regarding levels of risk factors and disease. What more is, we might not be able to extrapolate our findings to other ethnicities. According to a former study based on STR there was no difference in CVD risk and death when comparing twins with singletons [165]. In the fourth study the incidence rate for ischemic stroke was 6.8 per 1000 person-years (average age 70 years old). This is slightly higher than has been previously reported by a study on stroke incidence on cohorts based on Finnish and Swedish populations aged 60-69 years [166]. Moreover, an earlier study based on STR reported that the incidence rates of psychiatric illness were similar between twins and singletons [167].

7 Concluding remarks

Study I

The study emphasizes the importance of acknowledging the contribution of non-additive genetic effects to the risk of CVD. The results imply that dominant genetic models also should be considered when molecular studies are conducted in order to identify genes associated or linked to CVD blood biomarkers analyzed in the study. Moreover, unique environmental factors were shown to significantly impact all of the studied traits. The study also suggests that the equal environment assumption should be examined when conducting heritability studies on twin populations.

Study II

The study confirms that Lp-PLA_2 has a heritable component, although the environmental component contributing to trait variance was larger than what has previously been reported. It was also the first study to conduct variance partitioning on anti-PC levels, revealing both a genetic and unique environmental component underlying trait variance. Given that both of these biomarkers have previously been suggested to be associated with CVD risk and that anti-PC appears to be independent from other risk markers, further examination of Lp-PLA_2 and anti-PC's role as CVD biomarkers is necessitated. Genetic variants found to be associated with Lp-PLA_2 include genetic variants related to LDL, apoB, and TC levels, therefore the genetic correlation estimated in multivariate twin analyses could indicate genetic pleiotropy.

Study III

This large prospective population-based study indicates that depression is an important risk factor for later development of CVD. The association with depression was more pronounced for ischemic stroke risk, while neither antidepressants nor depression were associated with CHD. Nevertheless, since many previous studies have found a positive association between depression and CHD[75] replication of these non-significant findings in other studies is warranted. The association with CVD was strongest among depressed patients who have not used antidepressants. But the study design does not permit any conclusions on biological interactions, since for example the patients that did not consume antidepressants could have been affected by an additional set of risk factors than those who did use antidepressants. We also conclude that in general, use of antidepressants should not be utilized as a proxy measurement for depression. The study suggests that individuals who at some point in life are clinically depressed should be monitored more closely for their increased risk of developing CVD, particularly ischemic stroke. Further studies are needed to confirm and gain a deeper understanding of the association between clinical depression and stroke.

Study IV

In conclusion, evidence showing that clinical depression is a contributor to stroke development is increasing. The association was not accounted for by any of the covariates used in this study, including several important stroke risk factors. Studies conducting rigorous analyses on the link between depression and different stroke subtypes would be very valuable. Biological mechanisms underlying the observed association between depression and stroke have yet to be unraveled. We observed a marked difference in strength of association for self-reported depression compared to clinical depression. This indicates that extra caution is warranted when using self-reported index of depressive symptoms in epidemiological settings as a proxy for clinical depression. Prospective studies elucidating the biological mechanisms behind the relationship between depression and stroke are required, for not only would they enhance our biological understanding of stroke, but also provide clues about the mechanisms behind depression.

8 Future perspectives

8.1 Studies I and II

The last years, genome-wide data have revolutionized the ability to detect genetic risk variants underlying a range of biological traits. Nevertheless, in many cases these risk variants only explain a small proportion of the proposed heritability of the traits [168]. In this section novel ways on how to identify genetic risk variants underlying biological traits as well as how to estimate heritability will be discussed. In a paper by Yang et al. it was shown that common genetic variants do explain around half of the heritability for human height [169,170]. They showed this by applying a linear model to regress phenotypic similarity on genotypic similarity utilizing SNP data on a population consisting of unrelated individuals. Instead of inferring about additive relatedness from pedigree information (as is done in twin studies) this information is taken from SNP data, since the population consists of unrelated individuals the problem with shared environment was solved. They found that additive genetic effects explained 45% of the variance in height, which is much less than what has been found using classical twin design (80%), but much larger than had been discovered in genome-wide association studies (10%). The authors reasoned that the missing heritability of 35% was due to incomplete linkage disequilibrium (LD) between the SNPs in the chip and many of the actual causal genetic variants, and that individually these causal variants have small effects. In other words, genetic variants who are not in LD with SNPs in common chip arrays (rare genetic variants or variants in genomic regions poorly covered by the chip SNPs) with small effect sizes would explain the remaining heritability, however since they used an identity-by-state approach [171] on population with small sample size it was difficult to capture the contribution of many of the causal genetic variants. The authors concluded that prospective genome-wide studies of very large sample sizes will be required to detect genetic variants with very low minor allele frequency and small effect sizes.

A study by Zuk et al. provided a different explanation for the missing heritability, they concluded that the missing heritability is not missing at all, rather that genetic interactions (epistasis) create phantom heritability [172]. Their biological reasoning was that in many instances several biological inputs (e.g. proteins or chemical reactants) are required for a biological process to take place. They also referred to several studies on model organisms in which epistasis have been found [173,174]. They introduced a thinking framework on how this could be affecting heritability estimates by developing a limiting pathway (LP) model. Briefly, the LP model assumes that many biological traits depends on a number of rate-limiting inputs (k), these inputs themselves follow a standard additive model (the variance is explained by additive genetics and environmental factors). If k=1 then the LP model is additive, however if k is greater than 1 then according to the LP model this will result in phantom heritability, in other words the phantom heritability will increase as the number of k increases. The authors

rationalized that complex biological traits will more likely be affected by epistasis and consequently phantom heritability than traits of simpler biological architecture.

They proposed a model for heritability estimation based on unrelated individuals in recent genetically isolated population, isolated populations because then an identity-by-descent (IBD) [171] approach can be used. Hence, it would be more possible to capture the causal genetic variants with lower minor allele frequency. Simply put, with dense genotype data it could be possible to measure if the phenotypic similarity between individuals depends on genetic similarity (IBD sharing). They do not recommend methods based on relatives since they share alleles to a very large degree making it very difficult to tease apart epistasis from additive genetic effects. Moreover the levels of shared environment might vary between different constellations of relatives which further complicate the estimation of heritability. This model bears resemblance with the model utilized by Yang et al. [170], with the major difference being that it should be conducted on genetically isolated populations.

It remains to see if many more genetic variants will be found as study sample sizes increase and if indeed the genetic variance explained by genotypes will greatly increase. Or if the heritability (more specifically the additive heritability) measured by classical family design is inflated due to genetic interactions. Either way, developments of novel methods for estimating heritability, especially in an era where we have access to extensive molecular genetic data, should be promoted.

8.2 Studies III and IV

Many improvements can be made in observational studies investigating the association between depression and CVD. The construction of a clinical depression variable is least to say challenging. Many factors need to be considered, heterogeneity has been a major methodological limitation in studies III and IV, the specificity of the diagnoses was however weighed against sensitivity. The use of ICD-7 and ICD-8 codes brings with it a great deal of uncertainty of depression classification, a range of psychiatric conditions could particularly be included in ICD-7 where only one code is available for defining depressive mood. A possible solution would be to use a cohort consisting of a younger population and only confining to diagnostic codes from ICD-9 and ICD-10. Collection of medical records in which DSM-IV classifications are included would of course facilitate in obtaining a more detailed and accurate definition of depression.

To my knowledge, validation studies on unipolar depression diagnoses in the Swedish national patient register has yet to be performed. A study validating unipolar depression in the Danish Psychiatric Central Research Register found the validity to be satisfactory, around 75% of the patients who had received a single depressive episode in the register could also be diagnosed with single depressive episode according to Schedules for Clinical Assessment in Neuropsychiatry interview [175]. Extrapolating their results might be adequate due to ethnic/cultural similarities. Moreover, the psychiatric

care in Denmark and Sweden are to a large extent non-privatized, and the use of ICD-10 codes has been in practice since 1990's in both countries. Nonetheless, it would be prudent to carry out a study to assess the validity of unipolar depression diagnoses in the Swedish national patient register, and extend it to comprise diagnoses such as recurrent depressive disorder.

In future study designs it could be interesting to model the degree of depression severity, in order to capture a plausible dose-response relationship between depression and CVD. A former study demonstrated that the way ICD-10 grades unipolar depressive disorder (mild, moderate, severe) is clinically useful, since it predicts relapse and suicidal risk [176]. Variables such as admissions to inpatient versus outpatient hospital settings, the number of hospital admissions due to depression, and the age of depression onset could also be included in the models since they are indicative of depression severity. Inspection of comorbid conditions, by for instance thoroughly scanning through the additional ICD-coded diagnoses at each hospital admission could result in the finding of a major confounder. But considerations of these variables would require huge sample sizes and a real effort to reduce selection bias.

In study III, when investigating the relationship between antidepressants, depression and CVD, the strongest association with CVD was observed for depressed patients who did not use antidepressants [177]. This might have implied that associations between antidepressants and CVD observed in study III as well as previous studies could have been due to confounding by indication. The exact relationship between antidepressants and CVD is however debatable due to the very conflicting results presented in previous studies. One recent large population-based study suggested that depressed patients that used antidepressants with higher affinity for the serotonin transporter have a slightly elevated risk of stroke compared to users of antidepressants with lower affinity for the serotonin transporter [178]. This would be interesting to study more thoroughly in other study populations. A review by the Cochrane stroke review group based on 56 trials investigating the effect of SSRI on stroke recovery found that overall SSRIs appeared to improve disability, neurological impairment, and depression after stroke. They could not find any evidence that one type of SSRI was superior to another. They advised that efforts should be made on conducting prospective large-scale trials to assess whether SSRI indeed is beneficial for stroke recovery and should be given routinely to stroke patients [179].

Finally, access to high quality data for CVD ascertainment and health status such as neuroimaging data (CT and MRI scans), electrocardiography measurements, data from coronary/neurovascular angiography, duplex imaging of extracranial arteries, and measurements of physical activity would provide a higher accuracy of outcome and covariate definitions. It could however be fair to assume that severely depressed patients are less prone to take part in cohorts where a very active participation is necessary, thus register-based data might sometimes offer the best solution.

9 Acknowledgements

My PhD education has been an enthralling journey, and what I will remember most from the journey are all the faces I have encountered. I would like to take the opportunity to express my gratitude to;

First and foremost my main supervisor, the talented ***Patrik Magnusson****, for teaching me the core of scientific methodology, for giving me a free hand to experiment, for encouraging scientific skepticism, and last but not least for the killer jokes! You have been a huge support in my pursuit of knowledge and for that I will be ever grateful.*

My co-supervisor, ***Anna Bennet****, for assisting me in scientific writing and epidemiological thinking, and for all the delightful academic and non-academic conversations.*

My co-supervisor, ***Keith Humphreys****, for sharing your deep knowledge in biostatistics and for helping me out with the tricky methodological aspects of my studies.*

My co-supervisor, ***Erik Ingelsson****, for including me in your research group during the conference trips and for helping me out with the ICD-codes.*

Nancy Pedersen*, for guiding me through the immense amount of data available in STR and for teaching me how to write eloquently.*

Paul Lichtenstein, *for welcoming me into your research group.*

Henrik Grönberg*, for creating a very nice research environment at MEB.*

*Members of the Heart and Mind group****; Tove Fall, Andrea Ganna, He Gao, Lotte Gerritsen, Stefan Gustafsson, Marcel den Hoed, Sara Hägg, Ida Karlsson, Katherine Kasiman, Jitender Kumar, Kathleen Ma, Emil Rehnberg, Johanna Sieurin, Ci Song****, and* ***Fei Yang****, for all the interesting scientific discussions.*

My fourth floor peeps ***Ralf Kuja-Halkola*** *and* ***Therese Ljung,*** *for having shared my stress, headaches, and laughs. The camaraderie among us really helped me endure the PhD education. Ralf, thanks a million for tutoring me in biostatistics and for being an awesome friend.*

Adina Feldman, *you put so much effort into improving the PhD education at MEB. You really care about the well-being of other students, thanks for providing a gentle shoulder when needed, and for hosting fancy parties!* ***Tong Gong,*** *it has been a real pleasure to get to know you. Thanks for the intercultural exchange, for swimming and going to the gym with me, and for all the personal talks. We need to figure out what kind of sports activity to engage in next.* ***Kaavya Narasimhalu****, even though we are many miles apart we remain friends. It doesn't seem to bug you that I sometimes get lost in a philosophical maze, thanks for the conversations about anything and everything.*

I am very grateful that you joined me in my Icelandic escapade, you gals! I hope I will embark on new adventures together with you in the future.

Stephanie Bonn, Sara Christensen, Hatef Darabi, Rezin Dilshad, Miriam Elfström, Martin Fransson, Alexander Grankvist, Lovisa Högberg, Thomas Frisell, Robert Karlsson, Lisa Möller, Maria Sandberg, *and* ***Karin Sundström****, for your social company, interesting conversations and for all the party invitations!*

*Current and former students, staff and faculty at MEB****; Amy Leval, Anastasia Iliadou Nyman, Erik Pettersson, Giorgio Tettamanti, Simon Kyaga, Daniela Mariosa, Ove Strind, Edoardo Colzani, Rozita Broumandi, Anna Kähler, Vilhelmina Ullemar, Christina Hultman, Christina Persson, Zack Yusof, Ulrika Eriksson, Henrik Passmark, Henrik Larsson, Carolyn Cesta, Lennart Martinsson, Alexander Viktorin, Erika Nordenhagen, Roohi Rustum, Camilla Ahlqvist, Niklas Långström, Gunilla Sonnebring, Amir Sariaslan, Yudi Pawitan, Sara Ekberg, Michael Broms, Arvid Sjölander, Mun-Gwan Hong, Denny Rönnberg, Shuyang Yao, Kamila Czene, Hasse Walum, Emma Frans, Barbro Sandin, Pär Sparén,*** *and all others! Keep up the good spirit!*

All of those I have been lucky enough to befriend throughout the years, especially ***Any, Asrin, Emsada, Nadia, Tanzina, Teulip, toastmaster Ghazaleh, Lejla, Fousia, Rifat,*** *for the heart-to-hearts, fun and vivid memories deeply treasured, and long-lasting friendships.*

The members of ***Shapla,*** *for making sure that the Bangladeshi diaspora is thriving!*

My gazillion cousins all over the world, especially ***Sabrina, Sharmin, Shahrina, Natasha, Shakeel, Olivier, Lima, Lira, Blossom and Dervish,*** *for life-long bonding!*

My extended family, particularly ***Manna khala*** *and* ***khalu,*** *for being there for me at all times. You always keep your doors open and make me feel welcome.*

Nishu *and* ***Uwe,*** *for being awesome brother-in-laws.* ***Baby Isac*** *(nephew) for being such a champ! And rest of the in-laws, both in Sweden and elsewhere, for your kind spirit.*

My sisters; ***Dearest Sharlin,*** *it's been heartening to see you grow throughout the years, little sister, thanks for making me grow as a person.* ***Reefat,*** *your selfless ways and extremely considerate nature have constituted a tremendous resource in my life.* ***Rubaiyat,*** *you arrived in Sweden at 14 years of age, how did you manage to learn new things so swiftly? Your over-human working capacity is yet another source of wonderment. Thanks for being the supportive "big sister".*

***Dad**, "Where the mind is without fear and the head is held high*
Where knowledge is free
Where the world has not been broken up into fragments
By narrow domestic walls
Where words come out from the depth of truth
Where tireless striving stretches its arms towards perfection
Where the clear stream of reason has not lost its way
Into the dreary desert sand of dead habit
Where the mind is led forward by thee
Into ever-widening thought and action
Into that heaven of freedom, my Father, let my country awake" – Rabindranath Tagore

***Mum**, you are the root of my existence, wherever you are is my homeland.*

10 References

1 Mendis, S., Puska, P. & Norrving, B. Global atlas on cardiovascular disease prevention and control. (Geneva, 2011).

2 Madden, J. A. Role of the vascular endothelium and plaque in acute ischemic stroke. *Neurology* **79**, S58-62, (2012).

3 Ross, R. Atherosclerosis--an inflammatory disease. *N Engl J Med* **340**, 115-126, (1999).

4 Steinberg, D. Atherogenesis in perspective: hypercholesterolemia and inflammation as partners in crime. *Nat Med* **8**, 1211-1217 (2002).

5 Hansson, G. K. Inflammation, atherosclerosis, and coronary artery disease. *N Engl J Med* **352**, 1685-1695 (2005).

6 Charo, I. F. & Taub, R. Anti-inflammatory therapeutics for the treatment of atherosclerosis. *Nat Rev Drug Discov* **10**, 365-376 (2011).

7 Braunwald, E. Unstable angina: an etiologic approach to management. *Circulation* **98**, 2219-2222 (1998).

8 Wilson, P. W. *et al.* Prediction of coronary heart disease using risk factor categories. *Circulation* **97**, 1837-1847 (1998).

9 Di Angelantonio, E. *et al.* Lipid-related markers and cardiovascular disease prediction. *Jama* **307**, 2499-2506 (2012).

10 Kelly, T. N. *et al.* Systematic review: glucose control and cardiovascular disease in type 2 diabetes. *Ann Intern Med* **151**, 394-403 (2009).

11 Miller, M. *et al.* Triglycerides and cardiovascular disease: a scientific statement from the American Heart Association. *Circulation* **123**, 2292-2333 (2011).

12 Morrison, A. & Hokanson, J. E. The independent relationship between triglycerides and coronary heart disease. *Vasc Health Risk Manag* **5**, 89-95 (2009).

13 van 't Riet, E. *et al.* HbA1c is an independent predictor of non-fatal cardiovascular disease in a Caucasian population without diabetes: a 10-year follow-up of the Hoorn Study. *Eur J Prev Cardiol* **19**, 23-31 (2012).

14 Sarwar, N. *et al.* Triglycerides and the risk of coronary heart disease: 10,158 incident cases among 262,525 participants in 29 Western prospective studies. *Circulation* **115**, 450-458 (2007).

15 Liuzzo, G. *et al.* The prognostic value of C-reactive protein and serum amyloid a protein in severe unstable angina. *N Engl J Med* **331**, 417-424 (1994).

16 Ridker, P. M. High-sensitivity C-reactive protein and cardiovascular risk: rationale for screening and primary prevention. *Am J Cardiol* **92**, 17K-22K (2003).

17 Navab, M. *et al.* The oxidation hypothesis of atherogenesis: the role of oxidized phospholipids and HDL. *Journal of lipid research* **45**, 993-1007 (2004).

18 Kolodgie, F. D. *et al.* Lipoprotein-associated phospholipase A2 protein expression in the natural progression of human coronary atherosclerosis. *Arterioscler Thromb Vasc Biol* **26**, 2523-2529 (2006).

19 Packard, C. J. Lipoprotein-associated phospholipase A2 as a biomarker of coronary heart disease and a therapeutic target. *Curr Opin Cardiol* **24**, 358-363 (2009).

20 Hoffmann, M. M. *et al.* Genetic variants and haplotypes of lipoprotein associated phospholipase A2 and their influence on cardiovascular disease (The Ludwigshafen Risk and Cardiovascular Health Study). *J Thromb Haemost* **7**, 41-48 (2009).

21 Tsimikas, S. *et al.* Lipoprotein-associated phospholipase A2 activity, ferritin levels, metabolic syndrome, and 10-year cardiovascular and non-cardiovascular mortality: results from the Bruneck study. *Eur Heart J* **30**, 107-115 (2009).

22 Stremler, K. E., Stafforini, D. M., Prescott, S. M., Zimmerman, G. A. & McIntyre, T. M. An oxidized derivative of phosphatidylcholine is a substrate for the platelet-activating factor acetylhydrolase from human plasma. *J Biol Chem* **264**, 5331-5334 (1989).
23 Dada, N., Kim, N. W. & Wolfert, R. L. Lp-PLA2: an emerging biomarker of coronary heart disease. *Expert Rev Mol Diagn* **2**, 17-22 (2002).
24 Koenig, W. & Khuseyinova, N. Lipoprotein-associated and secretory phospholipase A2 in cardiovascular disease: the epidemiological evidence. *Cardiovasc Drugs Ther* **23**, 85-92 (2009).
25 Thompson, A. *et al.* Lipoprotein-associated phospholipase A(2) and risk of coronary disease, stroke, and mortality: collaborative analysis of 32 prospective studies. *Lancet* **375**, 1536-1544 (2010).
lipoprotein-associated phospholipase A2 ameliorates inflammation and decreases atherosclerotic plaque formation in ApoE-deficient mice. *PLoS One* **6**, e23425 (2011).
27 Wilensky, R. L. *et al.* Inhibition of lipoprotein-associated phospholipase A2 reduces complex coronary atherosclerotic plaque development. *Nat Med* **14**, 1059-1066 (2008).
28 Serruys, P. W. *et al.* Effects of the direct lipoprotein-associated phospholipase A(2) inhibitor darapladib on human coronary atherosclerotic plaque. *Circulation* **118**, 1172-1182 (2008).
29 Guerra, R. *et al.* Determinants of plasma platelet-activating factor acetylhydrolase: heritability and relationship to plasma lipoproteins. *J Lipid Res* **38**, 2281-2288 (1997).
30 Persson, M., Berglund, G., Nelson, J. J. & Hedblad, B. Lp-PLA2 activity and mass are associated with increased incidence of ischemic stroke: a population-based cohort study from Malmo, Sweden. *Atherosclerosis* **200**, 191-198 (2008).
31 Briles, D. E. *et al.* Antiphosphocholine antibodies found in normal mouse serum are protective against intravenous infection with type 3 streptococcus pneumoniae. *J Exp Med* **153**, 694-705 (1981).
32 Mold, C., Rodic-Polic, B. & Du Clos, T. W. Protection from Streptococcus pneumoniae infection by C-reactive protein and natural antibody requires complement but not Fc gamma receptors. *J Immunol* **168**, 6375-6381 (2002).
33 Padilla, N. D., Ciurana, C., van Oers, J., Ogilvie, A. C. & Hack, C. E. Levels of natural IgM antibodies against phosphorylcholine in healthy individuals and in patients undergoing isolated limb perfusion. *J Immunol Methods* **293**, 1-11 (2004).
34 de Faire, U. *et al.* Low levels of IgM antibodies to phosphorylcholine predict cardiovascular disease in 60-year old men: effects on uptake of oxidized LDL in macrophages as a potential mechanism. *J Autoimmun* **34**, 73-79 (2010).
35 Binder, C. J. *et al.* Pneumococcal vaccination decreases atherosclerotic lesion formation: molecular mimicry between Streptococcus pneumoniae and oxidized LDL. *Nat Med* **9**, 736-743 (2003).
36 Su, J. *et al.* Antibodies of IgM subclass to phosphorylcholine and oxidized LDL are protective factors for atherosclerosis in patients with hypertension. *Atherosclerosis* **188**, 160-166 (2006).
37 Sjoberg, B. G. *et al.* Low levels of IgM antibodies against phosphorylcholine-A potential risk marker for ischemic stroke in men. *Atherosclerosis* **203**, 528-532 (2009).

38 Frostegard, J. *et al.* Atheroprotective natural anti-phosphorylcholine antibodies of IgM subclass are decreased in Swedish controls as compared to non-westernized individuals from New Guinea. *Nutr Metab (Lond)* **4**, 7 (2007).
39 Rubin, E., Gorstein, F., Rubin, R., Schwarting, R. & Strayer, D. *Rubin's Pathology: Clinicopathologic Foundations of Medicine*. Fourth edn, 1437-1441 (Lippincott Williams & Wilkins, 2005).
40 Adams, H. P., Jr. *et al.* Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment. *Stroke* **24**, 35-41 (1993).
41 Bamford, J., Sandercock, P., Dennis, M., Burn, J. & Warlow, C. Classification and natural history of clinically identifiable subtypes of cerebral infarction. *Lancet* **337**, 1521-1526 (1991).
42 Goldstein, L. B. *et al.* Guidelines for the primary prevention of stroke: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke* **42**, 517-584 (2011).
43 O'Donnell, M. J. *et al.* Risk factors for ischaemic and intracerebral haemorrhagic stroke in 22 countries (the INTERSTROKE study): a case-control study. *Lancet* **376**, 112-123 (2010).
44 Djebali, S. *et al.* Landscape of transcription in human cells. *Nature* **489**, 101-108 (2012).
45 Dunham, I. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012).
46 De, S. Somatic mosaicism in healthy human tissues. *Trends Genet* **27**, 217-223 (2011).
47 Kazazian, H. H., Jr. Mobile elements: drivers of genome evolution. *Science* **303**, 1626-1632 (2004).
48 Stamatoyannopoulos, J. A. What does our genome encode? *Genome Res* **22**, 1602-1611 (2012).
49 Baillie, J. K. *et al.* Somatic retrotransposition alters the genetic landscape of the human brain. *Nature* **479**, 534-537 (2011).
50 Somel, M. *et al.* MicroRNA-driven developmental remodeling in the brain distinguishes humans from other primates. *PLoS Biol* **9**, e1001214 (2011).
51 Kimura, M. Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. *Nature* **267**, 275-276 (1977).
52 Mu, X. J., Lu, Z. J., Kong, Y., Lam, H. Y. & Gerstein, M. B. Analysis of genomic variation in non-coding elements using population-scale sequencing data from the 1000 Genomes Project. *Nucleic Acids Res* **39**, 7058-7076 (2011).
53 Sauna, Z. E. & Kimchi-Sarfaty, C. Understanding the contribution of synonymous mutations to human disease. *Nat Rev Genet* **12**, 683-691 (2011).
54 Prufer, K. *et al.* The bonobo genome compared with the chimpanzee and human genomes. *Nature* **486**, 527-531 (2012).
55 Britten, R. J. Divergence between samples of chimpanzee and human DNA sequences is 5%, counting indels. *Proc Natl Acad Sci U S A* **99**, 13633-13635 (2002).
56 Levy, S. *et al.* The diploid genome sequence of an individual human. *PLoS Biol* **5**, e254 (2007).
57 Mathers, C., Boerma, T. & Fat, D. M. in *The Global Burden Of Disease 2004 Update* Ch. 16, (World Health Organization, 2008).
58 Socialstyrelsen. Äldres behov av psykiatrisk vård och stöd. 20 (2012).

59 Nestler, E. J. *et al.* Neurobiology of depression. *Neuron* **34**, 13-25 (2002).
60 Manji, H. K., Drevets, W. C. & Charney, D. S. The cellular neurobiology of depression. *Nat Med* **7**, 541-547 (2001).
61 de Kloet, E. R., Joels, M. & Holsboer, F. Stress and the brain: from adaptation to disease. *Nat Rev Neurosci* **6**, 463-475 (2005).
62 Brown, E. S., Varghese, F. P. & McEwen, B. S. Association of depression with medical illness: does cortisol play a role? *Biol Psychiatry* **55**, 1-9 (2004).
63 Chourbaji, S. *et al.* IL-6 knockout mice exhibit resistance to stress-induced development of depression-like behaviors. *Neurobiol Dis* **23**, 587-594 (2006).
64 Raison, C. L., Capuron, L. & Miller, A. H. Cytokines sing the blues: inflammation and the pathogenesis of depression. *Trends Immunol* **27**, 24-31 (2006).
65 Haack, M. *et al.* Plasma levels of cytokines and soluble cytokine receptors in psychiatric patients upon hospital admission: effects of confounding factors and diagnosis. *J Psychiatr Res* **33**, 407-418 (1999).
66 Steptoe, A., Kunz-Ebrecht, S. R. & Owen, N. Lack of association between depressive symptoms and markers of immune and vascular inflammation in middle-aged men and women. *Psychol Med* **33**, 667-674 (2003).
67 Alexopoulos, G. S. *et al.* 'Vascular depression' hypothesis. *Arch Gen Psychiatry* **54**, 915-922 (1997).
68 Grool, A. M. *et al.* Lacunar Infarcts in Deep White Matter Are Associated with Higher and More Fluctuating Depressive Symptoms During Three Years Follow-up. *Biol Psychiatry* **73**, 169-176 (2013).
69 Mast, B. T. *et al.* Vascular disease and future risk of depressive symptomatology in older adults: findings from the Health, Aging, and Body Composition study. *Biol Psychiatry* **64**, 320-326 (2008).
70 Herrmann, L. L., Le Masurier, M. & Ebmeier, K. P. White matter hyperintensities in late life depression: a systematic review. *J Neurol Neurosurg Psychiatry* **79**, 619-624 (2008).
71 Avendano, M. *et al.* Socioeconomic status and stroke incidence in the US elderly: the role of risk factors in the EPESE study. *Stroke* **37**, 1368-1373 (2006).
72 Salaycik, K. J. *et al.* Depressive symptoms and risk of stroke: the Framingham Study. *Stroke* **38**, 16-21 (2007).
73 Hackett, M. L., Yapa, C., Parag, V. & Anderson, C. S. Frequency of depression after stroke: a systematic review of observational studies. *Stroke* **36**, 1330-1340 (2005).
74 Narushima, K., Kosier, J. T. & Robinson, R. G. A reappraisal of poststroke depression, intra- and inter-hemispheric lesion location using meta-analysis. *J Neuropsychiatry Clin Neurosci* **15**, 422-430 (2003).
75 Nicholson, A., Kuper, H. & Hemingway, H. Depression as an aetiologic and prognostic factor in coronary heart disease: a meta-analysis of 6362 events among 146 538 participants in 54 observational studies. *Eur Heart J* **27**, 2763-2774 (2006).
76 Pan, A., Sun, Q., Okereke, O. I., Rexrode, K. M. & Hu, F. B. Depression and risk of stroke morbidity and mortality: a meta-analysis and systematic review. *Jama* **306**, 1241-1249 (2011).
77 Whooley, M. A. *et al.* Depressive symptoms, health behaviors, and risk of cardiovascular events in patients with coronary heart disease. *Jama* **300**, 2379-2388 (2008).
78 Leonard, B. E. & Song, C. Stress and the immune system in the etiology of anxiety and depression. *Pharmacol Biochem Behav* **54**, 299-303 (1996).

79 Marano, G. *et al.* Depression and the cardiovascular system: increasing evidence of a link and therapeutic implications. *Expert Rev Cardiovasc Ther* **7**, 1123-1147 (2009).
80 Musselman, D. L., Evans, D. L. & Nemeroff, C. B. The relationship of depression to cardiovascular disease: epidemiology, biology, and treatment. *Arch Gen Psychiatry* **55**, 580-592 (1998).
81 Roose, S. P. Treatment of depression in patients with heart disease. *Biol Psychiatry* **54**, 262-268 (2003).
82 Kendler, K. S., Gardner, C. O., Fiske, A. & Gatz, M. Major depression and coronary artery disease in the Swedish twin registry: phenotypic, genetic, and environmental sources of comorbidity. *Arch Gen Psychiatry* **66**, 857-863 (2009).
83 Blanchette, C. M., Simoni-Wastila, L., Zuckerman, I. H. & Stuart, B. A secondary analysis of a duration response association between selective serotonin reuptake inhibitor use and the risk of acute myocardial infarction in the aging population. *Ann Epidemiol* **18**, 316-321 (2008).
84 Hamer, M., David Batty, G., Seldenrijk, A. & Kivimaki, M. Antidepressant medication use and future risk of cardiovascular disease: the Scottish Health Survey. *Eur Heart J* **32**, 437-442 (2011).
85 Smoller, J. W. *et al.* Antidepressant use and risk of incident cardiovascular morbidity and mortality among postmenopausal women in the Women's Health Initiative study. *Arch Intern Med* **169**, 2128-2139 (2009).
86 Tata, L. J. *et al.* General population based study of the impact of tricyclic and selective serotonin reuptake inhibitor antidepressants on the risk of acute myocardial infarction. *Heart* **91**, 465-471(2005).
87 Kimmel, S. E. *et al.* The effect of selective serotonin re-uptake inhibitors on the risk of myocardial infarction in a cohort of patients with depression. *Br J Clin Pharmacol* **72**, 514-517 (2011).
88 Monster, T. B., Johnsen, S. P., Olsen, M. L., McLaughlin, J. K. & Sorensen, H. T. Antidepressants and risk of first-time hospitalization for myocardial infarction: a population-based case-control study. *Am J Med* **117**, 732-737(2004).
89 Sauer, W. H., Berlin, J. A. & Kimmel, S. E. Selective serotonin reuptake inhibitors and myocardial infarction. *Circulation* **104**, 1894-1898 (2001).
90 Sauer, W. H., Berlin, J. A. & Kimmel, S. E. Effect of antidepressants and their relative affinity for the serotonin transporter on the risk of myocardial infarction. *Circulation* **108**, 32-36 (2003).
91 Meier, C. R., Schlienger, R. G. & Jick, H. Use of selective serotonin reuptake inhibitors and risk of developing first-time acute myocardial infarction. *Br J Clin Pharmacol* **52**, 179-184 (2001).
92 O'Connor, C. M. *et al.* Antidepressant use, depression, and survival in patients with heart failure. *Arch Intern Med* **168**, 2232-2237 (2008).
93 Lichtenstein, P. *et al.* The Swedish Twin Registry: a unique resource for clinical, epidemiological and genetic studies. *J Intern Med* **252**, 184-205 (2002).
94 Magnusson, P. K. *et al.* The Swedish Twin Registry: Establishment of a Biobank and Other Recent Developments. *Twin Res Hum Genet*, 1-13 (2012).
95 Pedersen, N. L., Lichtenstein, P. & Svedberg, P. The Swedish Twin Registry in the third millennium. *Twin Res* **5**, 427-432(2002).
96 Rahman, I. *et al.* Genetic dominance influences blood biomarker levels in a sample of 12,000 Swedish elderly twins. *Twin Res Hum Genet* **12**, 286-294(2009).

97 Ludvigsson, J. F., Otterblad-Olausson, P., Pettersson, B. U. & Ekbom, A. The Swedish personal identity number: possibilities and pitfalls in healthcare and medical research. *European journal of epidemiology* **24**, 659-667(2009).
98 Ludvigsson, J. F. *et al.* External review and validation of the Swedish national inpatient register. *BMC Public Health* **11**, 450(2011).
99 Dalman, C. & Wicks, S. Barn och vuxnas nyttjande av psykiatrisk vård. (Stockholm, 2000).
100 Socialstyrelsen. Dödsorsaker 2011. Report No. 2012-7-3, (2012).
101 Weitoft, G. R. *et al.* Fyra år med läkemedelsregistret. Report No. 2009-11-1, (2009).
102 Statistics, N. C. B. o. Psychiatric In-Patient Care 1977. (Stockholm, 1981).
103 Baranzini, S. E. *et al.* Genome, epigenome and RNA sequences of monozygotic twins discordant for multiple sclerosis. *Nature* **464**, 1351-1356 (2010).
104 Bruder, C. E. *et al.* Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variation profiles. *Am J Hum Genet* **82**, 763-771 (2008).
105 Plomin, R., DeFries, J. C., McClearn, G. E. & McGuffin, P. *Behavioral Genetics*. Fifth edn, 161 (Catherine Woods, 2008).
106 Falconer, D. S. & Mackay, T. F. C. *Introduction to Quantitative Genetics*. Fourth edn, 128 (Longman Group, 1996).
107 Falconer, D. S. & Mackay, T. F. C. *Introduction to Quantitative Genetics* Fourth edn, 172 (Longman Group, 1996).
108 Smith, E. N. & Kruglyak, L. Gene-environment interaction in yeast gene expression. *PLoS Biol* **6**, e83(2008).
109 Kohler, H. P., Behrman, J. R. & Schnittker, J. Social science methods for twins data: integrating causality, endowments, and heritability. *Biodemography Soc Biol* **57**, 88-141 (2011).
110 Kirkwood, B. R. & Sterne, J. A. C. *Essential Medical Statistics*. Second edn, 87-92 (Blackwell Science, 2003).
111 Pawitan, Y. *In All Likelihood: Statistical Modelling and Inference Using Likelihood*. 381 (Oxford University Press, 2001).
112 White, H. Maximum Likelihood Estimation of Misspecified Models. *Econometrica* **50**, 25 (1982).
113 Hanley, J. A., Negassa, A., Edwardes, M. D. & Forrester, J. E. Statistical analysis of correlated data using generalized estimating equations: an orientation. *Am J Epidemiol* **157**, 364-375 (2003).
114 Friedewald, W. T., Levy, R. I. & Fredrickson, D. S. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. *Clin Chem* **18**, 499-502 (1972).
115 Neale MC, B. S., Xie G, Maes HH *Mx: Statistical Modeling*. 7th edn, (2006).
116 Bayoumi, R. A. *et al.* Heritability of determinants of the metabolic syndrome among healthy Arabs of the Oman family study. *Obesity (Silver Spring)* **15**, 551-556 (2007).
117 Beekman, M. *et al.* Heritabilities of apolipoprotein and lipid levels in three countries. *Twin Res* **5**, 87-97, doi:10.1375/1369052022956 (2002).
118 Chen, T. J. *et al.* [Heritability of serum lipids and lipoproteins and its related factors in twins aged 5 to 19 years of China]. *Zhonghua Yu Fang Yi Xue Za Zhi* **38**, 237-239 (2004).

119 Chien, K. L. *et al.* Familial aggregation of metabolic syndrome among the Chinese: report from the Chin-Shan community family study. *Diabetes Res Clin Pract* **76**, 418-424 (2007).
120 de Oliveira, C. M., Pereira, A. C., de Andrade, M., Soler, J. M. & Krieger, J. E. Heritability of cardiovascular risk factors in a Brazilian population: Baependi Heart Study. *BMC Med Genet* **9**, 32 (2008).
121 Fenger, M., Schousboe, K., Sorensen, T. I. & Kyvik, K. O. Variance decomposition of apolipoproteins and lipids in Danish twins. *Atherosclerosis* **191**, 40-47(2007).
122 Goode, E. L., Cherny, S. S., Christian, J. C., Jarvik, G. P. & de Andrade, M. Heritability of longitudinal measures of body mass index and lipid and lipoprotein levels in aging twins. *Twin Res Hum Genet* **10**, 703-711 (2007).
123 Henneman, P. *et al.* Prevalence and heritability of the metabolic syndrome and its individual components in a Dutch isolate: the Erasmus Rucphen Family study. *J Med Genet* **45**, 572-577, doi:jmg.2008.058388 [pii]
10.1136/jmg.2008.058388 (2008).
124 Hokanson, J. E. *et al.* Pleiotropy and heterogeneity in the expression of atherogenic lipoproteins: the IRAS Family Study. *Hum Hered* **55**, 46-50(2003).
125 Iliadou, A., Lichtenstein, P., de Faire, U. & Pedersen, N. L. Variation in genetic and environmental influences in serum lipid and apolipoprotein levels across the lifespan in Swedish male and female twins. *Am J Med Genet* **102**, 48-58 (2001).
126 Isaacs, A. *et al.* Heritabilities, apolipoprotein E, and effects of inbreeding on plasma lipids in a genetically isolated population: the Erasmus Rucphen Family Study. *Eur J Epidemiol* **22**, 99-105 (2007).
127 McQueen, M. B., Bertram, L., Rimm, E. B., Blacker, D. & Santangelo, S. L. A QTL genome scan of the metabolic syndrome and its component traits. *BMC Genet* **4 Suppl 1**, S96 (2003).
128 Middelberg, R. P., Spector, T. D., Swaminathan, R. & Snieder, H. Genetic and environmental influences on lipids, lipoproteins, and apolipoproteins: effects of menopause. *Arterioscler Thromb Vasc Biol* **22**, 1142-1147 (2002).
129 Snieder, H., van Doornen, L. J. & Boomsma, D. I. Dissecting the genetic architecture of lipids, lipoproteins, and apolipoproteins: lessons from twin studies. *Arterioscler Thromb Vasc Biol* **19**, 2826-2834 (1999).
130 Souren, N. Y. *et al.* Anthropometry, carbohydrate and lipid metabolism in the East Flanders Prospective Twin Survey: heritabilities. *Diabetologia* **50**, 2107-2116 (2007).
131 Velasquez-Melendez, G., Parra, F. C., Gazzinelli, A., Williams-Blangero, S. & Correa-Oliveira, R. Genetic determinants of risk factors for cardiovascular disease in a population from rural Brazil. *Hum Biol* **79**, 179-190 (2007).
132 Teslovich, T. M. *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707-713 (2010).
133 Ninio, E. *et al.* Platelet-activating factor-acetylhydrolase and PAF-receptor gene haplotypes in relation to future cardiovascular event in patients with coronary artery disease. *Hum Mol Genet* **13**, 1341-1351 (2004).
134 Tselepis, A. D., Dentan, C., Karabina, S. A., Chapman, M. J. & Ninio, E. PAF-degrading acetylhydrolase is preferentially associated with dense LDL and VHDL-1 in human plasma. Catalytic characteristics and relation to the monocyte-derived enzyme. *Arterioscler Thromb Vasc Biol* **15**, 1764-1773 (1995).

135 Ryu, S. K. *et al.* Phospholipase A2 enzymes, high-dose atorvastatin, and prediction of ischemic events after acute coronary syndromes. *Circulation* **125**, 757-766 (2012).
136 Drenos, F. *et al.* Integrated associations of genotypes with multiple blood biomarkers linked to coronary heart disease risk. *Hum Mol Genet* **18**, 2305-2316 (2009).
137 Suchindran, S. *et al.* Genome-wide association study of Lp-PLA(2) activity and mass in the Framingham Heart Study. *PLoS Genet* **6**, e1000928 (2010).
138 Grallert, H. *et al.* Eight genetic loci associated with variation in lipoprotein-associated phospholipase A2 mass and activity and coronary heart disease: meta-analysis of genome-wide association studies from five community-based studies. *Eur Heart J* **33**, 238-251 (2012).
139 Vinson, A. *et al.* A pleiotropic QTL on 2p influences serum Lp-PLA2 activity and LDL cholesterol concentration in a baboon model for the genetics of atherosclerosis risk factors. *Atherosclerosis* **196**, 667-673(2008).
140 Lenzini, L. *et al.* A twin study of heritability of plasma lipoprotein-associated phospholipase A2 (Lp-PLA2) mass and activity. *Atherosclerosis* **205**, 181-185 (2009).
141 Kinoshita, K. & Honjo, T. Unique and unprecedented recombination mechanisms in class switching. *Curr Opin Immunol* **12**, 195-198 (2000).
142 Grobbee, D. E. & Hoes, A. W. Confounding and indication for treatment in evaluation of drug treatment for hypertension. *Bmj* **315**, 1151-1154 (1997).
143 Nabi, H. *et al.* Does depression predict coronary heart disease and cerebrovascular disease equally well? The Health and Social Support Prospective Cohort Study. *Int J Epidemiol* **39**, 1016-1024 (2010).
144 Wassertheil-Smoller, S. *et al.* Depression and cardiovascular sequelae in postmenopausal women. The Women's Health Initiative (WHI). *Arch Intern Med* **164**, 289-298, doi:10.1001/archinte.164.3.289 (2004).
145 Figgitt, D. P. & McClellan, K. J. Fluvoxamine. An updated review of its use in the management of adults with anxiety disorders. *Drugs* **60**, 925-954 (2000).
146 Wagstaff, A. J., Cheer, S. M., Matheson, A. J., Ormrod, D. & Goa, K. L. Paroxetine: an update of its use in psychiatric disorders in adults. *Drugs* **62**, 655-703 (2002).
147 Frisell, T., Oberg, S., Kuja-Halkola, R. & Sjolander, A. Sibling comparison designs: bias from non-shared confounders and measurement error. *Epidemiology* **23**, 713-720 (2012).
148 Rahman, I. *et al. Prospective study of clinical depression, self-reported depressive symptoms and their association with stroke* (2013).
149 Gellis, Z. D. Assessment of a brief CES-D measure for depression in homebound medically ill older adults. *J Gerontol Soc Work* **53**, 289-303 (2010).
150 Radloff, L. S. The CES-D Scale: A Self-Report Depression Scale for Research in the General Population. *Applied Psychological Measurement* **1**, 385-401 (1977).
151 Kawamura, T., Shioiri, T., Takahashi, K., Ozdemir, V. & Someya, T. Survival rate and causes of mortality in the elderly with depression: a 15-year prospective study of a Japanese community sample, the Matsunoyama-Niigata suicide prevention project. *J Investig Med* **55**, 106-114 (2007).
152 Witte, T. K., Timmons, K. A., Fink, E., Smith, A. R. & Joiner, T. E. Do major depressive disorder and dysthymic disorder confer differential risk for suicide? *J Affect Disord* **115**, 69-78 (2009).
153 Liebetrau, M., Steen, B. & Skoog, I. Depression as a risk factor for the incidence of first-ever stroke in 85-year-olds. *Stroke* **39**, 1960-1965 (2008).

154 Canan, F. *et al.* Association of mean Platelet volume with DSM-IV major depression in a large community-based population: The MELEN study. *J Psychiatr Res* **46**, 298-302 (2012).
155 De Meyer, S. F., Stoll, G., Wagner, D. D. & Kleinschnitz, C. von Willebrand factor: an emerging target in stroke therapy. *Stroke* **43**, 599-606 (2012).
156 Fujikawa, T., Yamawaki, S. & Touhouda, Y. Background factors and clinical symptoms of major depression with silent cerebral infarction. *Stroke* **25**, 798-801 (1994).
157 Vermeer, S. E., Koudstaal, P. J., Oudkerk, M., Hofman, A. & Breteler, M. M. Prevalence and risk factors of silent brain infarcts in the population-based Rotterdam Scan Study. *Stroke* **33**, 21-25 (2002).
158 Rothman, K. J., Greenland, S. & Lash, T. L. *Modern Epidemiology*. 3 rev ed edn, 130-134 (Lippincott Williams and Wilkins, 2008).
159 Danielsson, M., Gilliam, H. & Hernström, Ö. in *Folkhälsorapport 2009* (2009).
160 Lindsted, K. D., Fraser, G. E., Steinkohl, M. & Beeson, W. L. Healthy volunteer effect in a cohort study: temporal resolution in the Adventist Health Study. *J Clin Epidemiol* **49**, 783-790 (1996).
161 Taylor, A. The consequences of selective participation on behavioral-genetic findings: evidence from simulated and real data. *Twin Res* **7**, 485-504 (2004).
162 Rahman, I. *et al.* Genetic and environmental regulation of inflammatory CVD biomarkers Lp-PLA2 and IgM anti-PC. *Atherosclerosis* **218**, 117-122 (2011).
163 Cole, S. R. *et al.* Illustrating bias due to conditioning on a collider. *Int J Epidemiol* **39**, 417-420 (2010).
164 Jurek, A. M., Greenland, S., Maldonado, G. & Church, T. R. Proper interpretation of non-differential misclassification effects: expectations vs observations. *Int J Epidemiol* **34**, 680-687 (2005).
165 Oberg, S. *et al.* Twinship influence on morbidity and mortality across the lifespan. *Int J Epidemiol* **41**, 1002-1009 (2012).
166 Hyvarinen, M. *et al.* The difference between acute coronary heart disease and ischaemic stroke risk with regard to gender and age in Finnish and Swedish populations. *Int J Stroke* **5**, 152-156 (2010).
167 Kendler, K. S., Pedersen, N. L., Farahmand, B. Y. & Persson, P. G. The treated incidence of psychotic and affective illness in twins compared with population expectation: a study in the Swedish Twin and Psychiatric Registries. *Psychol Med* **26**, 1135-1144 (1996).
168 Lander, E. S. Initial impact of the sequencing of the human genome. *Nature* **470**, 187-197 (2011).
169 Visscher, P. M., Yang, J. & Goddard, M. E. A commentary on 'common SNPs explain a large proportion of the heritability for human height' by Yang et al. (2010). *Twin Res Hum Genet* **13**, 517-524 (2010).
170 Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* **42**, 565-569 (2010).
171 Stevens, E. L. *et al.* Inference of relationships in population data using identity-by-descent and identity-by-state. *PLoS Genet* **7**, e1002287 (2011).
172 Zuk, O., Hechter, E., Sunyaev, S. R. & Lander, E. S. The mystery of missing heritability: Genetic interactions create phantom heritability. *Proc Natl Acad Sci U S A* **109**, 1193-1198 (2012).

173 Brem, R. B., Storey, J. D., Whittle, J. & Kruglyak, L. Genetic interactions between polymorphisms that affect gene expression in yeast. *Nature* **436**, 701-703 (2005).
174 Khan, A. I., Dinh, D. M., Schneider, D., Lenski, R. E. & Cooper, T. F. Negative epistasis between beneficial mutations in an evolving bacterial population. *Science* **332**, 1193-119 (2011).
175 Bock, C., Bukh, J. D., Vinberg, M., Gether, U. & Kessing, L. V. Validity of the diagnosis of a single depressive episode in a case register. *Clin Pract Epidemiol Ment Health* **5**, 4, doi:10.1186/1745-0179-5-4 (2009).
176 Kessing, L. V. Severity of depressive episodes according to ICD-10: prediction of risk of relapse and suicide. *Br J Psychiatry* **184**, 153-156 (2004).
177 Rahman, I. *et al. Clinical depression, antidepressant use and risk of future cardiovascular disease* (2012).
178 Castro, V. M. *et al.* Incident user cohort study of risk for gastrointestinal bleed and stroke in individuals with major depressive disorder treated with antidepressants. *BMJ Open* **2**, e000544 (2012).
179 Mead, G. E. *et al.* Selective serotonin reuptake inhibitors (SSRIs) for stroke recovery. *Cochrane Database Syst Rev* **11**, CD009286 (2012).

11 Supplementary Tables

Table S1 *Age at use of antidepressant and stroke risk*

Exposure group	N antidepressant users	Hazard ratio (95% CI)	P-value
Antidepressants (≤50 years)	607	2.12 (1.30-3.46)	0.003
Antidepressants (>50 years)	5636	1.15 (1.01-1.30)	0.035

Adjusted for birth year, sex, educational level, smoking status, hypertension and BMI. If the first record of antidepressant use was found in the SALT interview, then age at SALT interview was used.

Table S2 *Age at depression diagnosis and stroke risk*

Exposure group	N depressed	Hazard ratio (95% CI)	P-value
Clinical depression (≤50 years)	835	1.83 (1.21-2.76)	0.004
Clinical depression (>50 years)	1032	1.62 (1.17-2.26)	0.004

Adjusted for birth year, sex, educational level, smoking status, hypertension and BMI

Table S3 *Association between SBP and depression adjusted for age and sex*

	Model 1			Model 2	
No cases	**OR (95% CI)**	P-value	**No cases**	**OR (95% CI)**	P-value
311	0.99 (0.99-1.00)	0.11	85	0.99 (0.97-0.99)	0.02

Model 1: Depression preceding SBP measurement (unit mmHg)
Model 2: SBP measurement (unit mmHG) preceding depression diagnosis